

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 30-08-2015		2. REPORT TYPE Ph.D. Dissertation		3. DATES COVERED (From - To) -	
4. TITLE AND SUBTITLE HYPOTHESIS TESTING USING SPATIALLY DEPENDENT HEAVY-TAILED MULTISENSOR DATA			5a. CONTRACT NUMBER W911NF-14-1-0339		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 611102		
6. AUTHORS Arun Subramanian			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Syracuse University Office of Research 113 Bowne Hall Syracuse, NY 13244 -1200			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 63915-CS.4		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The detection of spatially dependent heavy-tailed signals is considered in this dissertation. While the central limit theorem, and its implication of asymptotic normality of interacting random processes, is generally useful for the theoretical characterization of a wide variety of natural and man-made signals, sensor data from many different applications, in fact, are characterized by non-Gaussian distributions. A common characteristic observed in non-Gaussian data is the presence of heavy-tails or fat tails. For such data, the probability density function (p.d.f.) of extreme values decays at a slower than exponential rate, implying that extreme events occur with greater					
15. SUBJECT TERMS Copula theory, dependence modeling, detection, heavy-tailed signals, heterogeneous sensing, information fusion, sensor fusion.					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			Pramod Varshney
					19b. TELEPHONE NUMBER 315-443-1060

Report Title

HYPOTHESIS TESTING USING SPATIALLY DEPENDENT HEAVY-TAILED MULTISENSOR DATA

ABSTRACT

The detection of spatially dependent heavy-tailed signals is considered in this dissertation. While the central limit theorem, and its implication of asymptotic normality of interacting random processes, is generally useful for the theoretical characterization of a wide variety of natural and man-made signals, sensor data from many different applications, in fact, are characterized by non-Gaussian distributions. A common characteristic observed in non-Gaussian data is the presence of heavy-tails or fat tails. For such data, the probability density function (p.d.f.) of extreme values decay at a slower-than-exponential rate, implying that extreme events occur with greater probability. When these events are observed simultaneously by several sensors, their observations are also spatially dependent. In this dissertation, we develop the theory of detection for such data, obtained through heterogeneous sensors. In order to validate our theoretical results and proposed algorithms, we collect and analyze the behavior of indoor footstep data using a linear array of seismic sensors. We characterize the inter-sensor dependence using copula theory. Copulas are parametric functions which bind univariate p.d.f.s, to generate a valid joint p.d.f. We model the heavy-tailed data using the class of alpha-stable distributions. We consider a two-sided test in the Neyman-Pearson framework and present an asymptotic analysis of the generalized likelihood test (GLRT). Both, nested and non-nested models are considered in the analysis. We also use a likelihood maximization-based copula selection scheme as an integral part of the detection process. Since many types of copula functions are available in the literature, selecting the appropriate copula becomes an important component of the detection problem. The performance of the proposed scheme is evaluated numerically on simulated data, as well as using indoor seismic data. With appropriately selected models, our results demonstrate that a high probability of detection can be achieved for false alarm probabilities of the order of 10^{-4} .

These results, using dependent alpha-stable signals, are presented for a two-sensor case. We identify the computational challenges associated with dependent alpha-stable modeling and propose alternative schemes to extend the detector design to a multisensor (multivariate) setting. We use a hierarchical tree based approach, called vines, to model the multivariate copulas, i.e., model the spatial dependence between multiple sensors. The performance of the proposed detectors under the vine-based scheme are evaluated on the indoor footstep data, and significant improvement is observed when compared against the case when only two sensors are deployed. Some open research issues are identified and discussed.

ABSTRACT

The detection of spatially dependent heavy-tailed signals is considered in this dissertation. While the central limit theorem, and its implication of asymptotic normality of interacting random processes, is generally useful for the theoretical characterization of a wide variety of natural and man-made signals, sensor data from many different applications, in fact, are characterized by non-Gaussian distributions. A common characteristic observed in non-Gaussian data is the presence of heavy-tails or fat tails. For such data, the probability density function (p.d.f.) of extreme values decay at a slower-than-exponential rate, implying that extreme events occur with greater probability. When these events are observed simultaneously by several sensors, their observations are also spatially dependent. In this dissertation, we develop the theory of detection for such data, obtained through heterogeneous sensors. In order to validate our theoretical results and proposed algorithms, we collect and analyze the behavior of indoor footstep data using a linear array of seismic sensors. We characterize the inter-sensor dependence using copula theory. Copulas are parametric functions which bind univariate p.d.f.s, to generate a valid joint p.d.f.

We model the heavy-tailed data using the class of α -stable distributions. We consider a two-sided test in the Neyman-Pearson framework and present an asymptotic analysis of the generalized likelihood test (GLRT). Both, nested and non-nested models are considered in the analysis. We also use a likelihood maximization-based copula selection scheme as an integral part of the detection process. Since many types of copula functions are available in the literature, selecting the appropriate copula becomes an important component of the detection problem. The performance of the proposed scheme is evaluated numerically on simulated data, as well as using indoor seismic data. With appropriately selected models, our results demonstrate that a high probability of detection can be achieved for false alarm probabilities

of the order of 10^{-4} .

These results, using dependent α -stable signals, are presented for a two-sensor case. We identify the computational challenges associated with dependent α -stable modeling and propose alternative schemes to extend the detector design to a multisensor (multivariate) setting. We use a hierarchical tree based approach, called vines, to model the multivariate copulas, i.e., model the spatial dependence between multiple sensors. The performance of the proposed detectors under the vine-based scheme are evaluated on the indoor footstep data, and significant improvement is observed when compared against the case when only two sensors are deployed. Some open research issues are identified and discussed.

HYPOTHESIS TESTING USING SPATIALLY DEPENDENT HEAVY-TAILED MULTISENSOR DATA

By

Arun Subramanian

B.E., Bhavnagar University, 2003

M.S., Syracuse University, 2011

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
December 2014

UMI Number: 3670414

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3670414

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Acknowledgment

Research was sponsored by ARO grant W911NF-14-1-0339.

Copyright © 2014 Arun Subramanian

All rights reserved

ACKNOWLEDGMENTS

This dissertation was completed under the guidance of Prof. Pramod K. Varshney. I am thankful to him for being with me through every step of my doctoral studies. Throughout this process, his patience and continued encouragement to persist has been critical in motivating me to work through problems when I was just about ready to give up. I am grateful to my committee – Prof. Raja Velu, Prof. Yingbin Liang, Prof. Senem Velipasalar, Prof. Fritz H. Schlereth and Dr. Thyagaraju Damarla – for providing insightful comments and suggestions.

I was fortunate to have several motivating discussions with my colleagues at the Sensor Fusion Laboratory. They have always provided me with a fresh perspective and have been a constant voice of encouragement. They have been willing participants during data collection exercises. Discussions with Ashok Sundaresan, Satish Iyengar, Swarnendu Kar, Hao He, Sid Nadendla, Aditya Vempaty and Bhavya Kailkhura were particularly helpful.

Family and friends outside this academic circle have also been of tremendous help: they have been a voice of reason when I have been impulsive, and reminded me to be human and social when I was too self-absorbed. I am truly grateful for their unwavering support.

Finally, and most importantly, this thesis would not have been possible if it were not for the constant encouragement, faith and patience that my parents have shown since I left home. Their love and support has been unyielding, for which I shall be eternally grateful.

TABLE OF CONTENTS

Acknowledgments	v
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Statistical approach to information fusion	3
1.2 Dependence modeling	6
1.3 Heavy-tailed signals	7
1.4 Literature review	8
1.4.1 Dependence as covariance	9
1.4.2 Nonlinear dependence: nonparametric approach	12
1.4.3 Nonlinear dependence: copula-based approach	13
1.5 Contributions and organization	15
2 Indoor Seismic Data - Acquisition and Analysis	18
2.1 Sensor description and setup	18
2.2 Experiments	19
2.3 Preliminary data analysis	21
2.3.1 Nonlinearity analysis of observed data	21
2.3.2 Tail behavior of the seismic data	28

2.4	Other datasets	29
2.5	Summary	31
3	Statistical Dependence and Copula Theory	32
3.1	Bivariate statistical dependence	33
3.1.1	Positive and negative dependence	33
3.1.2	Measures of dependence	35
3.2	Copula theory	40
3.2.1	Summary of some copula functions	42
3.2.2	Copulas and measures of dependence	45
3.2.3	Tail dependence coefficients as a measure of extremal dependence . . .	47
3.3	Summary	48
4	Detection of Dependent Heavy-tailed Data	49
4.1	Introduction	50
4.2	Signal Model	52
4.2.1	Stable distributions	53
4.2.2	Dependent stable signals	55
4.3	The detection problem	56
4.3.1	Nested hypotheses or nested copula models	60
4.3.2	Non-nested models	64
4.4	Performance evaluation	71
4.4.1	Simulated examples	71
4.4.2	Computational considerations	78
4.4.3	Footstep Detection	80
4.5	Summary	86

5	Dependence Modeling for Detection Using Multiple Sensors	88
5.1	Problem Formulation	89
5.1.1	Complete ignorance	90
5.1.2	Approximate modeling	91
5.1.3	Nonparametric marginal estimation	92
5.2	Construction of Multivariate Copulas	93
5.2.1	Vines	93
5.2.2	Copula selection	96
5.2.3	Node ordering	98
5.3	Detection algorithm	100
5.4	Results	101
5.5	Summary	106
6	Detection of Footsteps from Outdoor Data	107
6.1	Data collection and preprocessing	107
6.2	Overview of the detector	109
6.3	Results	110
6.4	Conclusion	110
7	Summary and Future Directions	114
7.1	Future directions	116
7.1.1	Distributed detection with α -stable dependent observations	116
7.1.2	Sequential detection	116
7.1.3	Misspecified marginals.	118
7.1.4	Bootstrap-based detection for dependent observations	118
7.2	Some additional open problems	119
	References	120

LIST OF TABLES

2.1	Percentage of frames detected as nonlinear.	27
2.2	Comparison of values of $\bar{\phi}^{\text{rev}}$ with $[\text{SE}_{\bar{\phi}^{\text{rev}}}]$ for Footstep and Background data . .	28
4.1	Symbols and Notations	57
4.2	Library of copula functions	58
4.3	Summary statistics for parameter estimates of seismic data.	83
4.4	Percentage of copulas selected under each hypothesis	86

LIST OF FIGURES

1.1	Parallel topology for data fusion	6
2.1	The GS 20DX geophone.	20
2.2	Sensor setup in one of the buildings.	22
2.3	Time series of the background signal.	23
2.4	Time series of a footstep trial.	24
2.5	Test for nonlinearity.	26
2.6	Probability distribution of the background data	29
2.7	Probability distribution of the footstep data	30
3.1	Anscombe's quartet	39
4.1	ROC for Example 1: dependent $S\alpha S$ distributions.	73
4.2	ROC for Example 2: nearly normal distributions.	76
4.3	ROC for Example 3: skewed and asymmetrically dependent distributions.	77
4.4	Contour plot for f_X^1 from Example 3. The X and Y axes are the supports for the marginal densities of X_1 and X_2 , respectively.	78
4.5	Scatter plot of pilot background data	82
4.6	Histogram of α and γ values for the footsteps data.	84
4.7	ROC for Background vs. Footstep detection.	87
5.1	D-vine over 4 elements. Labels indicate the copula density evaluated at each tree in the vine.	95

5.2	ROC comparing the different MDL-based selection criteria.	102
5.3	ROC comparing the different detectors obtained from different marginal models. SIC is used for copula selection. CI: Complete Ignorance, AM: Approximate Modeling, NP: Nonparametric modeling.	103
5.4	ROC illustrating the benefit of using λ_U -based node ordering.	104
5.5	ROC comparing multivariate copula based detector to various bivariate copula based detectors. $\{S_p, S_q\}$ represent sensor pairs for $p, q = 1, 2, 3, 4$ and $p \neq q$. The x-axis is on a logarithmic scale to emphasize low P_F values.	105
6.1	ROCs for the ARL dataset for 1 person vs. background detection.	111
6.2	ROCs for the ARL dataset for 2 persons vs. background detection.	112
6.3	ROCs for the ARL dataset for 1 person leading an animal vs. background detection.	113

CHAPTER 1

INTRODUCTION

Our lives today are constantly aided and enriched by various types of sensors, which are deployed ubiquitously. They perform different roles, based on the context of their deployment. For example, as a part of modern mobile devices, we commonly find GPS information overlaid over image data, and this forms the basis of an augmented reality system. When deployed as a part of the different living spaces we occupy, sensors such as CO₂ and infrared modalities can be used indoors, at the front end of an energy-aware intelligent indoor environmental control system. Traffic cameras and GPS sensors can be used outdoors to assist drivers navigate busy rush-hour traffic.

In each of the above applications, sensors of different types, i.e., *heterogeneous* sensors, are used to make complex inferences about an underlying observed process. This is similar, in many ways, to how we, as humans, combine or *fuse* different streams of information originating from our sense organs. Over the past two decades, the field of information fusion has been extensively studied and researched. Although there exists a rich body of literature, the increasing complexity of systems as well as the vast diversity of applications require constant revision to existing technologies and continued research in this area.

In many inference applications, it is sufficient to deploy sensors, such as seismic or acoustic

modalities, which are capable of providing one dimensional time-series data. Sensing modalities such as video or infrared cameras have the ability to provide richer quality of information, but are either not practical to deploy, or have other constraints that do not permit their use in certain applications. For example, in an urban combat scenario, soldiers may require the surveillance of cleared buildings. For this application, the use of video cameras may either require a deployment and setup time which is not available, or there may exist critical areas that need monitoring but are occluded from a camera's field of view. When sensors are used for patient monitoring in hospitals, it is quite common to have situations where privacy concerns preclude the use of a video or similar imaging modality. In this dissertation, we are motivated by such applications and, in particular, develop appropriate theory, and for validation, apply it to the data obtained from seismic sensors deployed for indoor personnel monitoring.

The outcome of the information fusion process is, usually, some form of inference about the scene or *phenomenon* being observed. The phenomenon is context specific and, therefore, varies with the application being considered, e.g., personnel movement for surveillance, patient health in a health-care facility or habitability of the room for an indoor environment control application. The inference tasks could consist of detecting or estimating some parameters, such as locations or tracks, that provide information for situational awareness. The inferred parameters are a function of the specific model being considered, and emerge from the context/application under consideration.

Data from sensors typically exhibit information heterogeneity that can arise from a wide variety of causes. The sensors deployed in a given region of interest, in the most general setting, may consist of rather disparate and incommensurate modalities. Even sensors of the same modality may exhibit differences in their sensing ability, due to differences during manufacturing, quality control or the duration and location of their deployment. Since these sensors also observe different aspects of the same phenomenon, their observations are also dependent. The nature of this dependence can be quite complex and nonlinear, especially in cases where

the signal may propagate through a non-homogeneous medium. Additionally, the nature of the phenomenon, as well as the medium, can potentially result in non-Gaussian sensor measurements.

Fault tolerance and enhanced performance are key systemic advantages that result from fusing heterogeneous information sources because of the diversity, redundancy and increased coverage that they provide. As a consequence of heterogeneity, the quality and quantity of information provided by each sensing “modality”, which can potentially include human intelligence, varies with each source. In this sense, the words “sensor” and “node” are used interchangeably here and refer to any source of data. Note that while local observations and inferences from a group of heterogeneous sensors monitoring the same phenomenon may exhibit statistical dependence, they still provide different characterizations of the phenomenon under observation. Thus, the entire network does not fail as a result of one modality getting compromised. However, an accurate characterization of the inter-modality dependence is necessary for making reliable system-wide inference.

The above considerations are central to the ideas explored in this dissertation. We primarily investigate detection problems, from an information fusion perspective, when sensor observations are heterogeneous, dependent and heavy-tailed (non-Gaussian). Throughout this dissertation, we use footstep detection as an example application. For this we consider indoor seismic signals that we have collected using geophone sensors. In the following sections, we systematically introduce the main ideas related to information fusion using heterogeneous, dependent, heavy-tailed multisensor data.

1.1 Statistical approach to information fusion

The typical information fusion problem consists of a suite of networked or non-networked “sensors” that are deployed in a region of interest (ROI). The word “sensor” is used to include not only physical sensors, but any source capable of providing information based on its obser-

vations of a phenomenon occurring within the ROI. Therefore, local decision makers such as human agents are also considered to be sensors. Additionally, when monitoring a phenomenon of interest, the suite of sensors may consist of heterogeneous sensors.

The statistical approach to information fusion considers that a *fusion center* (FC) receives data from L sensors, where the data are characterized using a probabilistic model. The nature of the problem being considered, together with the model specification, determines the specific inference scheme employed by the fusion center. From each of the L sensors, the fusion center receives a sequence of N observations, x_{ij} , $i = 1, 2, \dots, L$, $j = 1, \dots, N$. Any inference process can use the data either sequentially, one observation at a time with an appropriate update rule, or take a one-shot approach, where one block $N \times L$ observations are used for inference. For the algorithms and techniques proposed in this dissertation, we consider a one-shot approach.

Each x_{ij} is a realization of the random variable, X_i . In this dissertation, we consider that the random variables are independent and identically distributed over the index j , but are dependent over the index i . These X_i , in the most general setting, may represent analog (unquantized) data, soft decisions or quantized data, or 1-bit local (hard) decisions. This notation, therefore accommodates raw sensor observations as well as data obtained after local or sensor-level processing.

When data are quantized locally to a 1-bit resolution, they often represent the case when sensors have additional processing capability to take local decisions. From the fusion perspective, this is also known as decision-level fusion. Local sensor-level processing which results in either a one-bit or M -bit output (with small M , i.e., coarse quantization) is often used in wireless sensor networks (WSN). A typical WSN is a network comprising power and bandwidth constrained sensors as the nodes of the network; these sensors transmit their hard or soft decisions to the fusion center through a wireless channel.

Feature-level fusion refers to the case when sensors have the computational resources for

complex signal processing, such that some descriptive features may be extracted from the raw data. Such features may include likelihood values, spectral coefficients, or the coefficients obtained from some other transform domain operation. Classification problems often employ such features which, when appropriately designed, provide a multidimensional basis for discriminating between the various classes under consideration. Feature extraction is also an effective way to process signals, which in their raw unprocessed form, may be incommensurate. When X_i represents unprocessed observations, such a fusion scheme is referred to as data-level fusion. This also represents the case where there is no decentralization of the decision-making process. Effectively, the fusion center is the *only* processing unit in the entire system.

Irrespective of the fusion levels or the nature of quantization, the design and analysis of a fusion method, from a statistical perspective, requires the probabilistic specification of X_i . The *sensor model* for the i -th sensor is, for analog data, the univariate probability density function (p.d.f.) or, for quantized data, the probability mass function (p.m.f.) of X_i . The sensors may be deployed in various spatial configurations or topologies. In this dissertation, we assume that the sensors send their observations to the fusion center, in parallel – without communicating with each other. This architecture is called the parallel fusion architecture in the distributed inference literature (see Fig. 1.1). Here, each sensor is depicted by a different shape: this is to indicate that the sensors could be of possibly different modalities.

The FC applies a *fusion rule*, which is a function defined on all X_i and determines a final decision or parameter value, based on the inference task considered. An optimal fusion rule typically maximizes a cost function, which is defined for the *entire* system. Note that in the preceding discussion, we do not specify the statistical nature of the sensor output; we only specify that all X_i are the input to the fusion center. That is, any distortion to the sensor's output (e.g., additive noise, fading, channel attenuation, etc.) are not modeled separately and are accommodated within X_i .

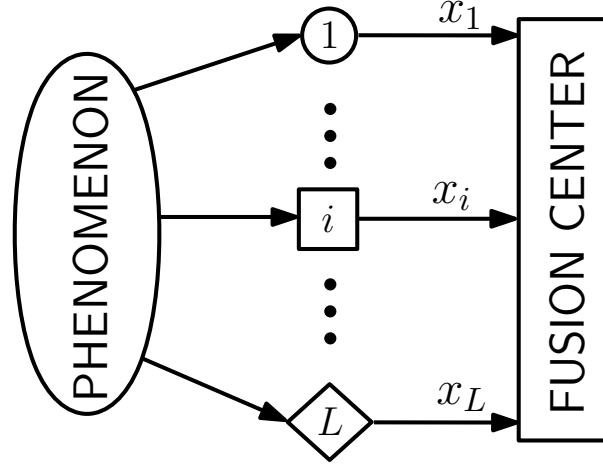


Fig. 1.1: Parallel topology for data fusion. Different shapes imply different sensor modalities.

1.2 Dependence modeling

When sensors observe a common phenomenon, as shown in Fig. 1.1, their measurements often exhibit spatial statistical dependence. This dependence may emerge in spite of sensors observing the phenomenon of interest as independent observers. For example, signals may be modeled as being embedded in additive *correlated* noise. Such a model is typically useful when sensors are deployed close to each other, e.g., in an acoustic array or a closely spaced array of antenna elements. When measurements are made using physical sensors, the relevant signals emerging from the source or phenomenon propagate through a *common* physical medium, before they are incident at the sensor. When the medium of propagation is non-homogeneous, the dependence structure between any X_i and $X_{i'}, i \neq i'$, can be significantly nonlinear. Hence, the commonly-used second-order measure – the correlation coefficient – becomes an inadequate measure of statistical dependence.

The issue of statistical dependence is even more complex when different sensor modalities are used. An observed phenomenon may give rise to disparate or incommensurate processes which are sensed and measured differently by modalities sensitive to the signals from those respective processes. For example, consider the phenomenon where acoustic and video modalities observe a person talking. Here, although the acoustic and video data are not coupled

via a shared medium of propagation, features extracted from voice data (acoustic sensor) and image sequences of lip movements (video sensor) are statistically dependent. In this example, dependence is induced by the phenomenon. A multimodal deployment necessarily implies heterogeneity. Suppose, for $1 \leq i \neq i' \leq L$, the p.d.f.s of X_i and $X_{i'}$ are denoted as f_{X_i} and $f_{X_{i'}}$, sensors i and i' are *heterogeneous* if $f_{X_i} \neq f_{X_{i'}}$. Note that if sensor i is an acoustic modality and i' provides video data, f_{X_i} and $f_{X_{i'}}$ may not be defined on the same support. This is an additional layer of complexity when modeling the joint p.d.f. $f_{X_i, X_{i'}}$. The joint distribution of sensor measurements is necessary for any inference task. In this dissertation, copulas, discussed in detail in Chapter 3, are used to construct valid joint distributions describing possibly nonlinear dependence structures, such that each X_i can be heterogeneous.

1.3 Heavy-tailed signals

Many important stochastic phenomena cannot be adequately modeled with distributions that decay exponentially in the tail. For such phenomena, extreme value measurements occur at a significantly greater frequency than is attributable to distributions that decay exponentially in the tail. One can typically observe a “spiky” signature in a time series plot of these measurements and such signals are often said to be *fat-tailed* or *heavy-tailed*.

Examples of such signals can be seen in applications such as finance, geology, climatology and bioengineering. In many scenarios arising from these applications, the detection of significant deviations or anomalies from a process describing a null hypothesis is an important task. Many of these anomalies can be characterized as extreme-value deviations from the null process, i.e., the anomalies occur with low probability and fall in the tail regions of the null hypothesis distributions. These problems have been studied in detail when the underlying distributions are well-behaved and easy to characterize. However, when distribution tails decay at slower-than-exponential rates, the inference task becomes difficult because of modeling and associated tractability issues.

When the observed data have heavy tails in their distribution, the consequences of improper model selection become more severe. An anomalous process being observed by multiple sensors such that observations are dependent implies that, effectively, the fusion center observes extreme co-movements in the distribution tails. Such events are called tail-dependent events. Development and selection of models capable of capturing this tail-dependence, also called extremal dependence, becomes an important component of the overall inference problem. Extremal dependence is especially relevant in the context of modern portfolio theory. When there does not exist sufficient diversity within a portfolio, the associated risk increases given an expected return or profit. When distributions characterizing the associated risk do not capture the tail-heaviness or tail-dependence, the likelihood corresponding to high risk values are underestimated, which affects the reliability of decisions. The application of improper models, where tail-dependence was inadequately quantified, was considered to be one of the causes for the financial crisis of 2007-2008.

In this dissertation, we focus on footstep signals, acquired from an array of seismic sensors, and show that they can be modeled as dependent heavy tailed signals. Using these signals as a motivating example, the theory for modeling and detecting spatially dependent heavy tailed signals is studied. The effect of model selection, as an integral part of the detection framework, is also analyzed.

1.4 Literature review

Multisensor signal processing may be viewed as a subset of the broader field of information fusion. Centralized formulations, where raw observations are available at the processing unit or fusion center, for several inference tasks are well known and available in standard textbooks [12, 47, 94]. Distributed inference, on the other hand, relies on the availability of a network that can either transmit local inferences/quantized measurements to the fusion center or arrive at a consensus solution by locally exchanging compressed/quantized information.

While research in this area has forked in various directions, the problems addressed can be categorized as either distributed detection [98] or decentralized estimation (e.g., see [66, 71, 78] and references cited therein).

This section reviews recent progress that has taken place in the field of multisensor signal processing, and focuses on developments where dependence information plays a significant role in the design. The aim of the discussion, as presented, is to motivate the relevance of our research presented in this dissertation. One of the major themes explored in this dissertation is the concept of statistical dependence. Therefore, this section discusses the literature in the context of different types of dependence models that have been employed over the years. The emphasis on dependence notwithstanding, the literature is quite extensive, and instead of being exhaustive, we concentrate on highlighting newer developments.

1.4.1 Dependence as covariance

Modeling dependence as a covariance matrix (or equivalently a correlation matrix) is arguably one of the most popular ways of characterizing dependence. It defines the dependence of jointly normal random variables and describes the linear dependence between random variables that possess a finite second moment. Due to the inherent simplicity associated with the use of second order statistics such as the correlation coefficient, it has been applied in various contexts in both centralized and distributed inference schemes.

Centralized schemes for correlated sensor observations

In the centralized paradigm, covariance-based dependence modeling is used extensively to model the dependency information for array signal processing applications, especially where it is reasonable to assume linearity of the medium of signal propagation. The most recent applications where these concepts of array signal processing have been applied are MIMO radar [59] and joint blind source separation (JBSS) [4], among others. In MIMO radar, several antenna

elements are used to transmit multiple probing signals that may be correlated or uncorrelated with one another. While traditional blind source separation problems are formulated using a single dataset, JBSS formulations are useful when analyzing multiple datasets as a group. An example of this is separating speech and audio signals in multiple frequency bands.

The fusion of EEG with fMRI data for the detection of schizophrenia is discussed by Correa et al. [19] where the brain tissue is modeled as a mixing channel, and hence the information fusion problem is posed as a JBSS problem and is solved using an approach based on multivariate canonical correlation analysis [48]. Canonical correlation analysis (CCA) is a technique which transforms the data matrix in such a way that it maximizes the amount of correlation between the entities exhibiting statistical dependence. It has also been used for audio-video fusion: Slaney and Covell [86] use CCA to measure the synchrony between acoustic features and video frames, while Kidron et al. [49] consider a CCA based approach to determine pixels in images that exhibit maximal correlation with the acquired audio signal.

Distributed inference using correlated data

Optimal schemes for distributed inference with correlated observations has also been a topic of considerable interest. In the case of distributed detection, it has been shown that the likelihood ratio based quantizer, which was optimal under the assumption of conditional independence, is no longer optimal when correlation is taken into account. Examples of the consequent loss in performance are provided by Aalo and Viswanathan [1]. In fact, earlier work by Tsitsiklis and Athans [97] has shown that the distributed detection problem with dependent observations is NP-complete. One way to get past the computational intractability is to assume some prior knowledge about the joint statistics: Drakopolous and Lee [25] examine the fusion rule for distributed detection under dependence by considering that the correlation coefficient is known, whereas Kam et al. [44] use the Bahadur-Lazarsfeld expansion of probability density functions.

Willett et al. [107] study the problem of distributed detection of a mean shift in corre-

lated Gaussian noise and establish how the nature of correlation affects the optimum fusion rule. They conclude that even for a simple two-sensor and linear correlation formulation the distributed detection problem “exhibits apparently very complicated behavior.” For this mean shift in correlated Gaussian noise problem, local quantizers designed using the likelihood ratio test (LRT) are, in general, not optimal. Willet et al. show that determining the parameter regions where this optimality may hold is itself a challenging task: while the optimality of the LRT can be determined for certain parameter regions, the problem is mostly intractable for other regions. Chen et al. [16] have recently proposed a more general formulation this problem. They introduce a hidden variable that induces conditional independence among the sensor observations so that many more distributed detection problems with dependent observations become tractable. This new framework allows for the identification of several classes of distributed detection problems with dependent observations whose optimal decision rules resemble the ones for the conditionally independent case. The new framework induces a decoupling effect on the forms of the optimal local decision rules for these problems, much in the same way as the conditionally independent case. This is in sharp contrast to the general dependent case where the coupling of the forms of local sensor decision rules often renders the problem intractable. Such decoupling enables the use of, for example, the person-by-person optimization approach to find optimal local decision rules. The two cases of distributed detection, deterministic signal in dependent noise, and detection of a random signal in independent noise, have become tractable under this new framework.

The decentralized estimation problem with correlated observations has been studied by Fang and Li [30]. They consider a power constrained wireless sensor network [92] and examine power allocation for spatially correlated sensor observations. Each sensor transmits a possibly nonlinear function of the parameter of interest, θ , that is corrupted by additive, correlated Gaussian noise. Bandwidth constrained formulations requiring quantized transmissions to the fusion center are also considered by Ribeiro and Giannakis [78]. However, they con-

sider a linear observation model, with θ being deterministic but unknown, and hence the sensor observations are conditionally independent. Krasnopeev et al. [52] present a distributed estimation scheme for the problem $x_i = \theta + n_i$, where x_i is the measurement of sensor i and the noise $\mathbf{n} = [n_i]$ is a multivariate Gaussian random vector which is correlated spatially across sensors. The covariance is assumed to be known at the fusion center. We note that all these problems are considered to be distributed since each local sensor transmits some local estimate of θ , which in its simplest form is the noise corrupted parameter itself. These formulations do not consider local, inter-node communication; the implications of this local communication aspect have been recently investigated by Kar et al. [46].

1.4.2 Nonlinear dependence: nonparametric approach

Nonparametric approaches to multisensor signal processing have been very popular in applications where it is infeasible to model *a priori* the complex dependencies that may exist between the signals/features acquired by the sensors. These methods, in essence, estimate or learn the joint distribution across sensor measurements directly from the data.

Machine learning techniques fall under this framework and are applicable largely when it is feasible to control environment variables in such a way that a representative training dataset may be collected. While this is apparently a stringent requirement, often with some preprocessing, a significant amount of information can be extracted from sensor observations. This has led to the successful application of machine-learning techniques for a wide variety of problems. Learning based methodologies have been successfully applied to multibiometric systems [11, 79]. Multibiometric systems achieve superior personnel identification performance by fusing information from two or more biometric modalities. The learning-based approach has also been popular for solving several object classification tasks [43, 64] and have traditionally focused on security and surveillance applications [58, 111]. Recently, challenges unique to emerging technologies such as ubiquitous and human-centered computing have led to new

research in areas such as object tracking and affect recognition [108, 109].

When viewed from an information fusion perspective, nonparametric designs offer tangible advantages over methods described in Section 1.4.1. Fusion of heterogeneous or multimodal information is possible since disparate modalities are not constrained to a multivariate normal approximation. For example, Butz and Thiran [14] use the mutual information and joint entropy between audio and video data as a measure of dependence; the joint density required for the computation of these quantities is estimated from the data using the nonparametric Parzen's estimator [102]. Graphical models such as Bayesian networks generalize hidden Markov models and have also been successfully used for audio-visual tracking [8, 22, 43]. Algorithms for distributed fusion using graphical models have been developed by Çetin et al. [15].

1.4.3 Nonlinear dependence: copula-based approach

As indicated earlier, we employ copulas to characterize joint distributions. Copulas are parametric functions that couple univariate marginal distribution functions to the corresponding multivariate distribution function. A copula-based formulation is attractive because the spatial correlation among sensor observations can get manifested in several different, potentially non-linear ways and many families of copula functions have been specified in the literature to address this issue. Further, while nonparametric formulations are known to converge to the true distribution asymptotically, they also suffer from scalability issues stemming from the curse of dimensionality. Recently, considerable progress has been made in the study of copulas and their applications in statistics. The usage of copulas is widespread in the fields of econometrics and finance [17] and they are beginning to be used in the signal and image processing context [23, 39, 63, 89].

In the fusion context, the use of copulas can be first found, in the operations research context, in a paper by Jouini and Clemen [42]. They propose a copula-based method for the aggregation of expert opinions. They take a Bayesian approach in their formulation: they con-

sider Bayesian decision-makers who make subjective assessments about the observed process. This subjective assessment is encoded as a univariate marginal distribution, which is combined, along with the assessments of other experts using a copula. In order to elicit the copula parameter, they propose a multivariate extension of Kendall's tau as a measure of dependence between multiple experts.

Sundaresan et al. [88] first considered the case of distributed detection for dependent observations, using a copula based framework. They derived the optimum fusion rules for a Neyman-Pearson detector. In their work, they found that the fusion rules under copula-based dependence have a similar form as the Bahadur-Lazarsfeld expansions, proposed by Kam et al. Sundaresan and Varshney [87] also design and analyze the performance of a copula-based estimation scheme for the localization of a radiation source.

Iyengar et al. [36] have investigated the general framework of copula-based detection of a phenomenon being observed jointly by heterogeneous sensors. They quantify the performance loss due to copula misspecification and demonstrate that a detector using a copula selection scheme based on area under the receiver operating characteristic (ROC) can provide significant improvement over models assuming independence. Their results on a NIST multibiometric dataset show that the copula based approach is versatile and can fuse not only heterogeneous sensor measurements, but can also be applied to fuse different algorithms. The tractability issue of fusing dependent quantized data is addressed by Iyengar et al. [37]. In this paper, the authors found that injecting a suitably designed noise variable, the optimum fusion rule can be approximated for a minimum level of distortion. The problem of intractability, due to the presence of multiple coupled integrals, is reduced to a problem of multiplying characteristic functions, similar to the way in which frequency selective filtering is done.

1.5 Contributions and organization

The main contributions of the research results presented in this dissertation to the signal processing and information fusion literature, are as follows:

- A data-collection procedure was designed and executed to create a dataset of footstep signals obtained using seismic sensors. The data can be used for data-driven problem-specific tasks such as investigating procedures for indoor personnel occupancy detection and activity classification. It also serves as an example of heavy-tailed data exhibiting spatio-temporal dependence.
- The theory of detection for dependent α -stable signals is studied using a copula-based approach for dependence characterization. Issues such as model nesting and model selection are studied in depth. We derive the necessary and sufficient conditions for multivariate model nesting using a copula-based approach for distribution modeling. We also derive asymptotic results for the probabilities of false-alarm and detection, for both nested and non-nested hypotheses for detecting dependent heavy tailed observations.
- A vine-based approach is proposed for modeling multisensor dependence. Using bivariate copula building blocks, the vine-based approach allows us to construct multivariate models free of symmetry constraints. The effect of model selection and node ordering is investigated in the context of footstep detection. A tail-dependence motivated algorithm is presented for establishing a node order for the base tree in the vine.

In Chapter 2, we discuss the data-collection process. The data was collected from a linear array of geophone seismic sensors. This chapter describes the sensor and data-acquisition hardware and also the data collection procedure. The heavy-tailed nature of the footstep data provides the motivation for considering detection schemes tailored specifically for dependent heavy-tailed data.

Chapter 3 explores the background on statistical dependence and introduces copula theory. Measures of dependence, other than the correlation coefficient, are surveyed and their connections to copula functions are also summarized.

Chapter 4 examines the problem of detection of dependent α -stable signals. We use the class of α -stable distributions to characterize the heavy-tailed nature of these signals. For typical applications, sensors make simultaneous measurements of a given phenomenon, and hence these heavy-tailed realizations are dependent across sensors. The inter-sensor dependence is modeled using copulas. We consider a two-sided test in the Neyman-Pearson framework and present an asymptotic analysis of the generalized likelihood test (GLRT). Both, nested and non-nested models are considered in the analysis. The performance of the proposed scheme is evaluated numerically on simulated data, as well as the indoor seismic data described in Chapter 2. With appropriately selected models, our results demonstrate that a high probability of detection can be achieved for false alarm probabilities of the order of 10^{-4} . While the theory presented in this chapter is valid for multiple-sensor deployments, we consider a two-sensor case for ease of exposition.

In Chapter 5, we address copula construction and model selection issues for the multi-sensor (i.e., multivariate) case. Using the Neyman-Pearson approach, we show that accounting for multivariate dependence leads to significant improvement over a bivariate approach, within the copula framework. The tree-based technique of *vines* are used for modeling the dependence across multiple sensors. The vine based approach is able to model asymmetric dependence between sensor observations.

Chapter 6 discusses the results obtained when the copula-based detection scheme is applied on outdoor data. The outdoor data was provided by the U. S. Army Research Laboratory (ARL) and was collected close to the southwest US border. The chapter discusses the results obtained, for footstep detection, when seismic data was fused with acoustic data.

Chapter 7 summarizes the salient concepts explored in this dissertation and examines directions for future research for copula-based inference.

CHAPTER 2

INDOOR SEISMIC DATA - ACQUISITION AND ANALYSIS

As indicated in Chapter 1, the research considered in this dissertation is motivated by applications where, due to various considerations, only one-dimensional signals are available, such as seismic or acoustic signals. In order to obtain a dataset that is representative of such application scenarios, we collected seismic data by deploying an array of geophone sensors in a typical indoor office environment. The performance of the proposed detectors, in Chapter 4 and Chapter 5, will be evaluated on the data thus collected. This chapter discusses the physical characteristics of the sensors and the data collection hardware in Section 2.1, and a description of the experiments for collecting the background and footstep data is provided in Section 2.2.

2.1 Sensor description and setup

Six GS 20DX geophones were used for the experiments. The electrical details of a typical sensor [31] are depicted in Figure 2.1(b) and the frequency response curve is depicted in Figure 2.1(c). Transduction is achieved by means of a moving coil over a magnetic core. The geo-

phones are designed to be floor mounted. Floor to sensor contact was achieved by means of a coupling bolt screwed to the sensor, which was held to the floor by means of a tripod base. This was done since tight coupling was not feasible as it requires structural penetration by means of a probe.

The sensors constitute a wired suite and are connected to a data-acquisition system (DAQ). The DAQ is essentially an AD converter with preset amplification and low level software for device control. The DAQ used for these experiments was a model PDL-MF: a PCI data-acquisition device developed by United Electronic Industries, Walpole, MA. The data-acquisition card provides for 8 analog input channels with an overall sampling rate of 50,000 samples/s and 16-bit quantization. At its maximum preset amplification factor of 10 it can faithfully (i.e., without clipping) digitize a signal of amplitude $\pm 1V$. While lower amplifications can accommodate a wider range of signal amplitude, this setting was selected as footsteps generate a voltage swing, in each sensor, of the order of only a few mV.

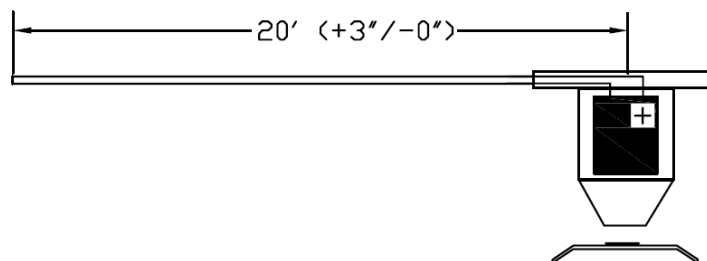
The DAQ was programmed, using a C++ library provided by the manufacturer, to acquire data at 5kHz. The DAQ was programmed on a non-real-time operating system (Microsoft Windows XP Professional) and therefore, file I/O operations for a 5kHz sampling rate is a challenging task. Sequential read-write execution leads to the read buffer in the AD converter getting filled up before the acquired data can be written, leading to data loss. This problem exists in spite of the manufacturer providing a large capacity circular buffer. The problem was solved by programming the read and write operations to execute as parallel threads.

2.2 Experiments

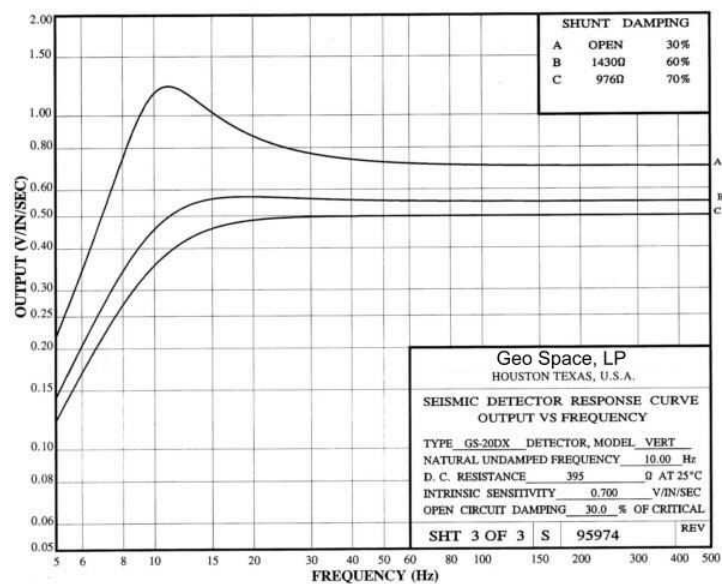
The six sensors were configured as a linear array. They were placed along the long edge of a hallway (see Figure 2.2). Data was collected in two (different) building hallways of similar construction. The sensors were placed along the long edge of the hallway. The distance between adjacent sensors was maintained at 5ft. The rationale for selecting a sampling rate of



(a)



(b)



(c)

Fig. 2.1: The GS 20DX geophone. (a) Sensor as housed and packaged (b) Electrical details: cable length and sensor polarity (c) Frequency response curve.

5kHz was so that, if necessary, high frequency information in the footstep data [29] could be utilized for detection/classification. However, considering the frequency response curve of the GS 20DX (see Fig. 2.1(c)) and the typical quasi-periodicity of footstep signals, the raw signal was uniformly down-sampled to 1024 samples/second.

Background data was collected by leaving the sensors in an isolated environment. Background data is approximately of a 4 minute duration. Multiple persons participated in the footstep data collection. The footstep data collected consists of 120 single-person trials (i.e., a given trial has exactly one participant walking along the hallway) and 120 two-person trials (a given trial has exactly two participants walking along the hallway). Each dataset consists of 60 trials from Building 1 and 60 trials from Building 2. The approximate duration of the data collected per trial is 12 seconds. The background and footstep signal from a single person trial are graphed in Fig. 2.3 and Fig. 2.4 respectively. In the following section, some analysis on data collected is presented, the analysis focuses on the presence of nonlinearity and heavy tailed behavior of the seismic data.

2.3 Preliminary data analysis

In this section, we present an analysis of the data. The nonlinear nature of the data is first explained. Signal nonlinearity, within the footsteps context, is strongly suggestive of a nonlinear mixing medium of signal propagation. The tail behavior of the data is analyzed next, and we demonstrate that the footstep data cannot be explained by a Gaussian or exponential tail-decay model.

2.3.1 Nonlinearity analysis of observed data

A signal $y(t)$ is said to be nonlinear if its current value cannot be predicted or expressed as a linear function of its past values. Let i denote the sensor index, i.e., $i = 1, 2, \dots, 6$. Each



Fig. 2.2: Sensor setup in one of the buildings.

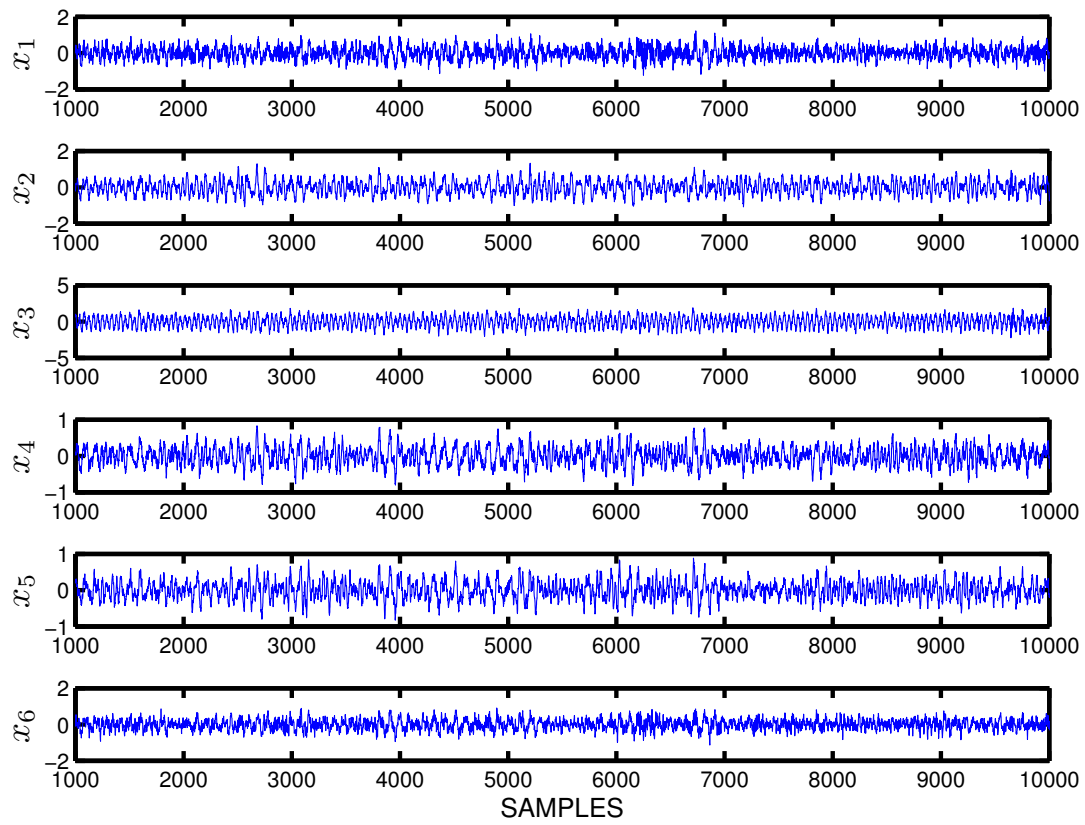


Fig. 2.3: Time series of the background signal.

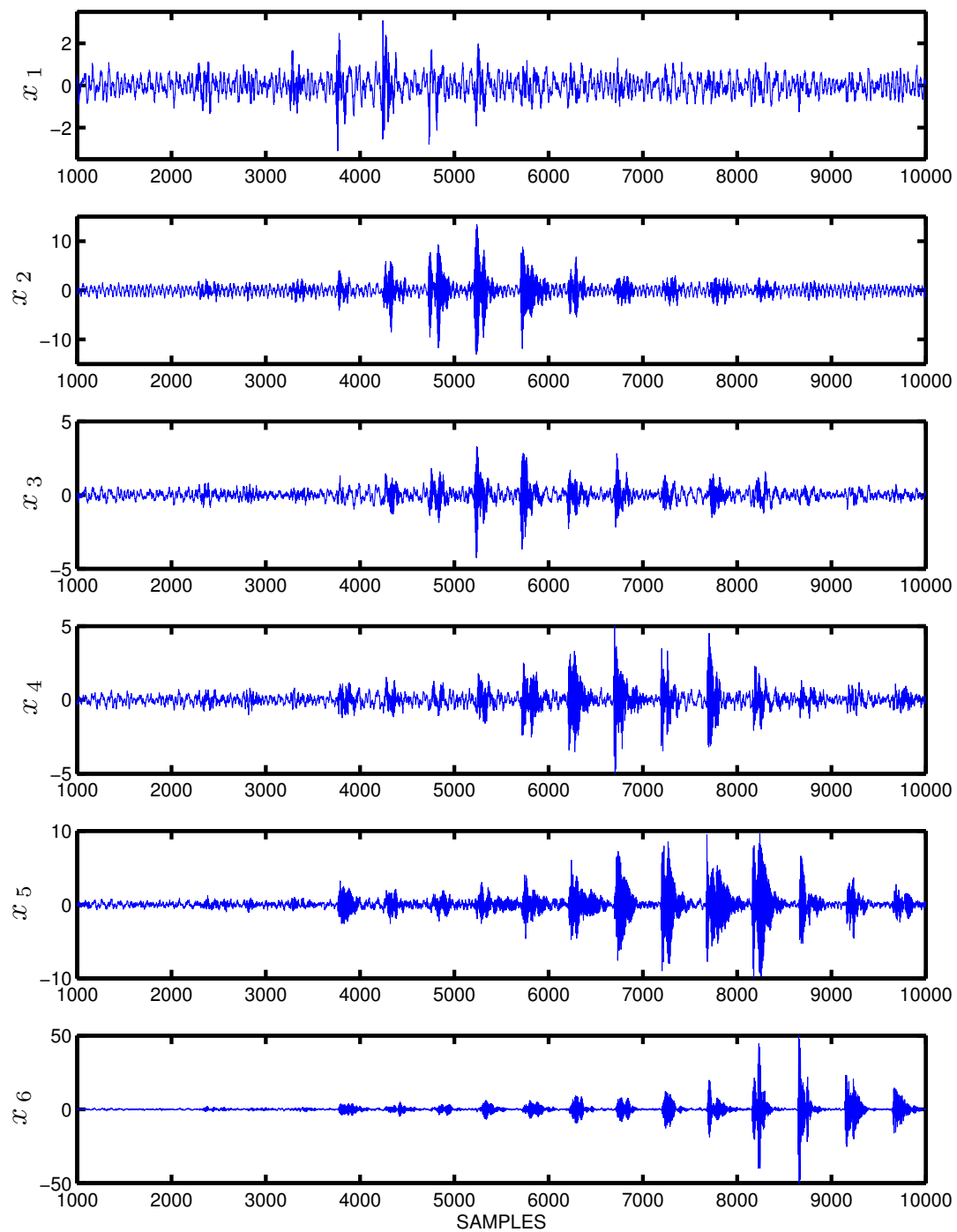


Fig. 2.4: Time series of a footstep trial. Nonstationarity and the impulsive nature of the signal is evident.

sensor observation, $y_i(t)$, is uniformly down-sampled to 1024 Hz. Each $y_i(t)$ is divided into 1 second overlapping frames. Denote by \mathcal{T}_r , the set of all time instants contained in the r th frame. Therefore, the cardinality of \mathcal{T}_r , $|\mathcal{T}_r|$, is 1024. The inter-frame overlap was set to 50%.

The method of surrogate data [82] is used to analyze the acquired seismic time series for the presence of nonlinearity. The null hypothesis states that the original time series is a realization of a linear Gaussian process (or monotonic transforms thereof). The idea is to generate a set of time series (surrogate data set) by resampling from the original measurements so that linear statistical properties of the original data are preserved in the surrogate data set. These surrogates are then, in essence, samples from a population consistent with the null hypothesis of linearity and can be used to estimate the distribution of a test statistic that can discriminate between the null (linearity) and alternative (nonlinearity). This statistic is computed for both the surrogate data and the original time series. If the statistic computed on the original time-series lies (significantly) in the tail of the distribution of the statistic corresponding to the surrogate, the null is rejected.

Following Schreiber [82], a third order statistic,

$$\phi^{\text{rev}} = \frac{1}{N-1} \sum_{n=2}^N (y[n] - y[n-1])^3, \quad (2.1)$$

is used to test for nonlinearity in our analysis. Here $y[n]$ is the sampled version of $y(t)$ and N is the number of samples in the r -th frame. A known property of a linear Gaussian process is that its statistics are symmetric under time reversal [103]; ϕ^{rev} measures the asymmetry of a series under time-reversal [82].

Each frame of $y_i(t)$ (for all i) is tested for the presence of nonlinearity as follows. Forty surrogates, $s_k(t) : k = 1, 2, \dots, 40$, are generated from a given frame of the original time series $y_i(t)$ using the iterative amplitude adjusted Fourier transform (IAAFT) [82]. The IAAFT algorithm is based on the amplitude adjusted Fourier transform (AAFT) algorithm [93], which samples from a normal distribution and the sampled sequence is ranked and scaled so that the

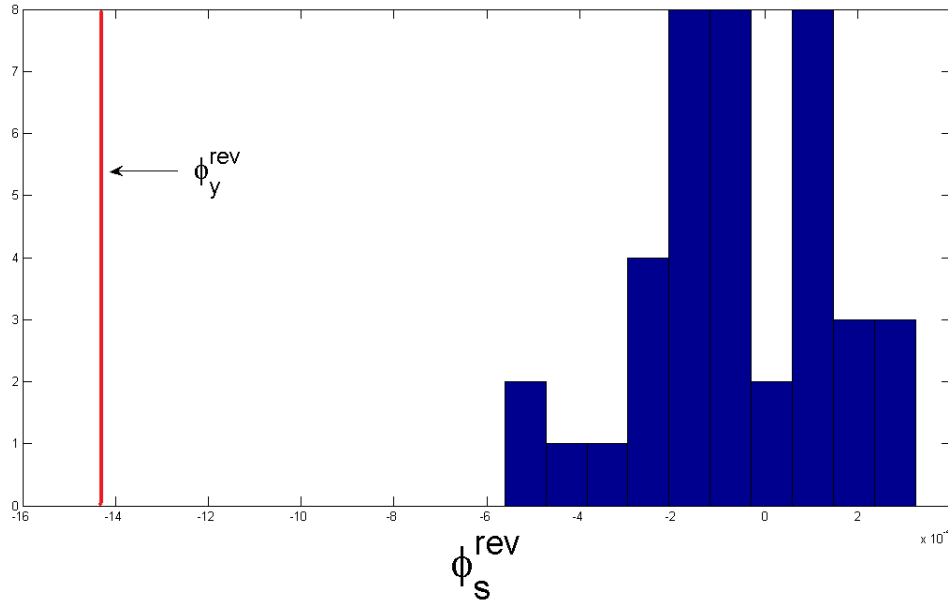


Fig. 2.5: Test for nonlinearity. Histogram is generated using the surrogate data. The statistic of the original time series is represented by the solid line labeled ϕ_y^{rev} .

amplitude spectra of the surrogates matches that of the frame under test while randomizing the phase uniformly between 0 and 2π . Schreiber notes that the AAFT algorithm is correct only *asymptotically*: it generates surrogates that are linear *and* have the same amplitude probability distribution (APD) as the original time-series as $N \rightarrow \infty$. The IAAFT algorithm, proposed in [83], iterates between amplitude adjustment and phase randomization until the surrogates and the original data have the same APD. The statistic in Eq. (2.1) is then computed for both the surrogates and the test data. For example, consider Fig. 2.5. The solid line indicates the value of ϕ_y^{rev} , the statistic computed for sensor data corresponding to a frame from the walking trials. The histogram of ϕ_s^{rev} , computed for the corresponding surrogates is also shown. It is evident that the footstep signal has a nonlinear structure to it as ϕ_y^{rev} does not lie within the distribution of the null hypothesis corresponding to linearity. Thus, the null hypothesis can be rejected.

Each frame of the footstep signal is tested for the presence of nonlinearity at 0.05 significance level ($\alpha = 0.05$) using a rank order test proposed by Theiler et al. (see Section 2.1, [93]).

Table 2.1: Percentage of frames detected as nonlinear.

<i>Sensor</i> <i>i</i>	<i>Footstep data</i>	
	<i>1 s frame</i>	<i>2 s frame</i>
1	25	20
2	19	26
3	12	14
4	11	11
5	17	20
6	19	24

The case when the frames are two seconds in duration is also considered and results are summarized in Table 2.1. We observe that a significant proportion of the walking frames are detected as nonlinear. However, the exact nature of the nonlinearity is not known and is difficult to ascertain. This, is also a different characterization of the nonstationarity present in the data, and therefore, motivates the use of semiparametric methods of inference with such data. We also compared the values of ϕ^{rev} obtained for the footstep and background data. The standard tests for normality, such as the Jarque-Bera test, confirm that the background data are normally distributed. After standardizing the footstep and background time-series data, values of ϕ^{rev} are computed over 1s and 2s frames. In Table 2.2, $\bar{\phi}^{\text{rev}}$, the mean value of ϕ^{rev} over the total number of frames, along with the standard error ($\text{SE}_{\bar{\phi}^{\text{rev}}}$) are shown for both frame durations. Similar to what we observe in Fig. 2.5, the numbers reveal that for a linear Gaussian process, values of $\bar{\phi}^{\text{rev}}$ are close to zero with narrow standard errors, implying that the values of ϕ^{rev} are spread about a narrow interval centered about 0. Values of ϕ^{rev} for the footstep data, on the other hand, lie significantly outside this region and are almost an order of magnitude greater than the typical $\bar{\phi}^{\text{rev}}$ values for linear Gaussian processes. However, since $\bar{\phi}^{\text{rev}}$ estimates for footstep data also possess larger standard errors, several time-series frames are classified as “linear” as seen in Table 2.1.

Table 2.2: Comparison of values of $\bar{\phi}^{\text{rev}}$ with $[\text{SE}_{\bar{\phi}^{\text{rev}}}]$ for Footstep and Background data

<i>Sensor</i> <i>i</i>	<i>Background</i>		<i>Footstep</i>	
	1 s <i>frame</i>	2 s <i>frame</i>	1 s <i>frame</i>	2 s <i>frame</i>
1	−0.0050 [12.67 · 10 ^{−5}]	−0.0035 [5.4 · 10 ^{−5}]	−0.0185 [6.1 · 10 ^{−4}]	−0.0115 [3.5 · 10 ^{−4}]
2	0.0005 [1.6 · 10 ^{−5}]	0.0004 [0.8 · 10 ^{−5}]	0.1557 [4.6 · 10 ^{−3}]	0.0938 [2.7 · 10 ^{−3}]
3	≈ −10 ^{−5} [2.2 · 10 ^{−5}]	≈ −10 ^{−5} [1.3 · 10 ^{−5}]	0.0120 [4.3 · 10 ^{−4}]	0.0139 [2.9 · 10 ^{−4}]
4	−0.0010 [2.7 · 10 ^{−5}]	−0.0013 [1.8 · 10 ^{−5}]	0.0205 [7.2 · 10 ^{−4}]	0.0248 [4.9 · 10 ^{−4}]
5	−0.0003 [2.3 · 10 ^{−5}]	−0.0003 [1.3 · 10 ^{−5}]	0.0147 [6.3 · 10 ^{−4}]	0.0176 [4 · 10 ^{−4}]
6	−0.0080 [7.6 · 10 ^{−5}]	−0.0084 [6.1 · 10 ^{−5}]	0.0056 [5.2 · 10 ^{−4}]	0.0079 [3.1 · 10 ^{−4}]

2.3.2 Tail behavior of the seismic data

We have also analyzed the collected seismic data for the tail behavior. The background data show the presence of exponential tails, and the footstep data show the presence of heavy tails which decay at a polynomial rate. An example of this tail behavior is shown in Fig. 2.6 and Fig. 2.7. We observe that not only do the tails of the background data have exponential decay, but they do this at a slightly sub-Gaussian rate. This can be explained considering the physical nature of the geophone, which damps sudden (discontinuous) excursions of the signal. An idea of this behavior can also be inferred from the frequency plot of the sensor in Fig. 2.1, which shows a damped response in the high-frequency regime. On the other hand, the heavy-tailed behavior of the footstep data is clearly visible in Fig. 2.7. In fact, we can even infer a polynomial decay in the tails of the footstep data. Note that for any distribution decaying at a polynomial rate in the tails, i.e., as $|x|^{-\alpha-1}$, the logarithm of this, i.e., $-(\alpha + 1) \log |x|$ will saturate for extreme values of x . This is precisely the behavior we observe in the p.d.f. plot

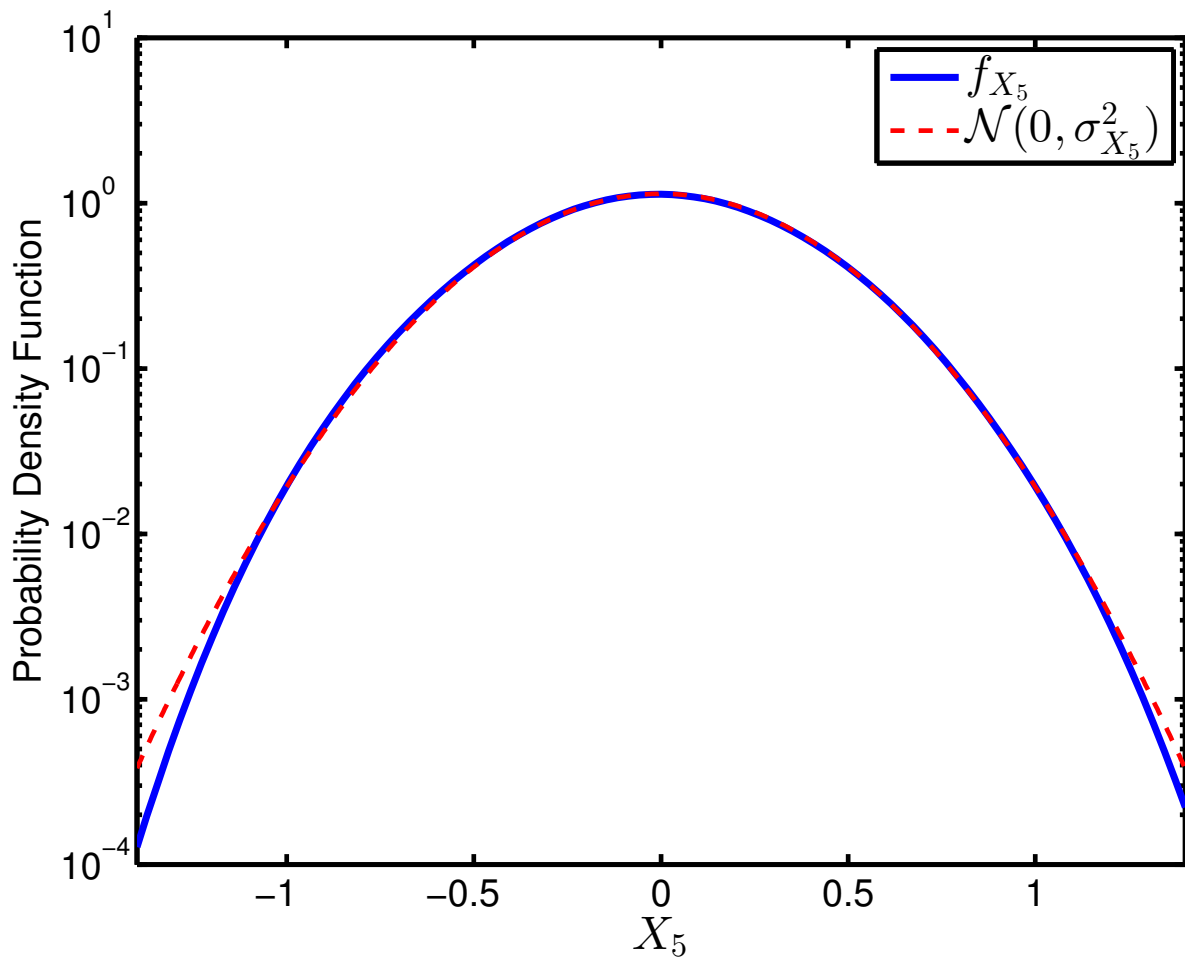


Fig. 2.6: Probability distribution of the background data from Sensor 5, compared to the p.d.f. of normal distribution with the same second-order moments as the background data. The Y -axis is plotted on a logarithmic scale.

for footstep data, where the Y -axis is plotted on a logarithmic scale. The significance of a tail decay rate of $|x|^{-\alpha-1}$ is explained in Chapter 4.

2.4 Other datasets

The research in this dissertation took place in collaboration with US Army Research Laboratory (ARL). In this effort, we also collected data using the unattended ground sensor (UGS) suite at ARL, at Adelphi, MD. Additional data, for an outdoor scenario, was collected by ARL near the

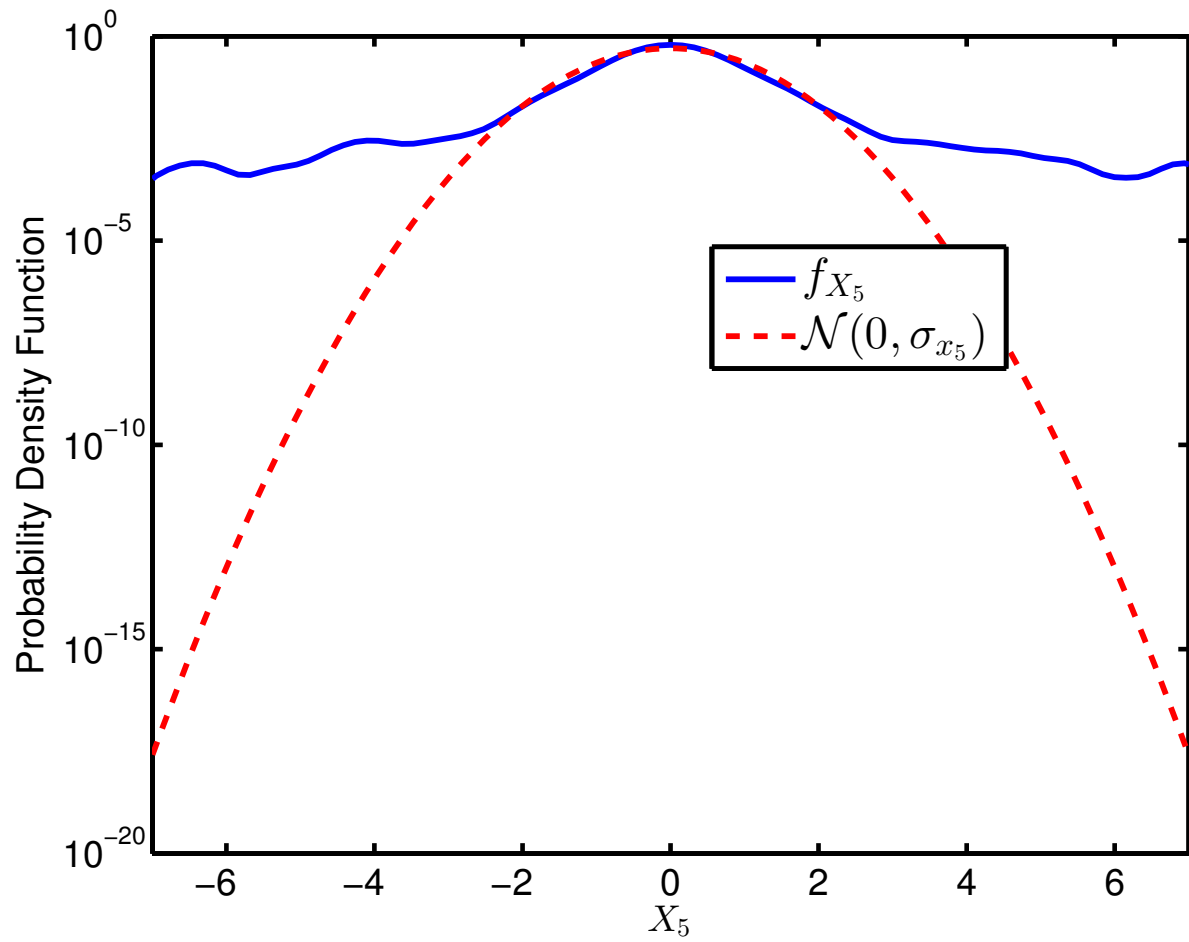


Fig. 2.7: Probability distribution of the footstep data from Sensor 5, compared to the p.d.f. of normal distribution with the same second-order moments as the footstep data. The Y-axis is plotted on a logarithmic scale.

southwest US border. The copula-based methods, proposed in Chapters 4 and 5, have also been applied to these datasets. For the purpose of demonstrating our detection methodology on real sensor data, in this dissertation, we focus on the results obtained using the dataset described in this chapter. However, we have applied similar methods to the indoor and outdoor ARL datasets and results based on these data are discussed in Chapter 6.

2.5 Summary

In this chapter we analyzed the nature of seismic data collected using geophone sensors in an indoor environment. The analysis revealed that the data corresponding to footstep activity exhibits temporal nonlinearity, with heavy-tailed behavior. The background data, on the other hand, are approximately normal. Time series plots of the footstep data also reveal that the data are spatially dependent, but signal nonlinearity will imply that the statistical dependence exhibited by the data will not be explainable by simple models. A more sophisticated understanding of statistical dependence is required, and appropriate models must be used for any sort of inference done using such data. While we have demonstrated the existence of complex spatio-temporal behavior using footstep data, such signal characteristics can be seen in other types of data too. The analyses presented in this chapter, therefore, motivate our research approach in this dissertation. We address the general theory of detecting such spatially dependent heavy-tailed data, and return to the footstep data example to apply our proposed methods.

CHAPTER 3

STATISTICAL DEPENDENCE AND

COPULA THEORY

Chapter 1 reviewed the recent research on signal processing for stochastically dependent observations. As noted in Section 1.4, parametric, semi-parametric and non-parametric techniques of dependence characterization have been extensively studied, and they find utility in a variety of applications. As a consequence, research that includes the consideration of dependence in various disciplines such as machine learning, information theory, speech processing, finance, and aerospace, among others, has led to a rich body of literature. Dependence modeling, in this dissertation, is based on copula theory, which can be categorized as either a parametric or semi-parametric approach to dependence modeling, depending on the formulation being considered. In this chapter, concepts and measures of dependence are discussed (Section 3.1), followed by an overview of copula theory (Section 3.2).

3.1 Bivariate statistical dependence

The topic of stochastic dependence has been studied extensively since Karl Pearson first defined the product-moment correlation. This section discusses several concepts and measures of bivariate dependence that have since sought to generalize Pearson's correlation coefficient. The focus on bivariate dependence is due to the fact that many concepts of multivariate dependence do not carry over as a simple extension of the bivariate case. Further, when exploring the idea of multivariate dependence, we use a pairwise scheme, in Chapter 5, based on the concept of *vines*. The topics covered here summarize a more detailed treatment of dependence concepts by Balakrishnan and Lai (see [7], Chapter 3 and Chapter 4). The discussion that follows in the next section will show how a copula-based characterization of joint distributions relates to these generalized descriptions of dependence.

3.1.1 Positive and negative dependence

For two continuous random variables, X and Y , *positive dependence* implies that large/small values of Y tend to accompany large/small values of X . In contrast, *negative dependence* implies that large/small values of Y tend to accompany small/large values of X . We discuss only concepts that are derived from positive dependence, since the negative dependence counterparts are analogous. Further, if the pair (X, Y) has a positive dependence, then $(X, -Y)$ has negative dependence on \mathbb{R}^2 . If there exists a constraint of positivity, $(X, 1 - Y)$ has negative dependence on the unit square. An important point to note is that while one may define positive dependence for the multivariate case, negative dependence is no more a mirror reflection of positive dependence. Six basic conditions describing positive dependence have been discussed in the literature [55]. These are enumerated below in the increasing order of stringency.

1. **Positive correlation.** Defined for positive linear correlation, i.e., $\text{cov}(X, Y) \geq 0$.
2. **Positive quadrant dependence (PQD).** $\mathbb{P}(X > x, Y > y) \geq \mathbb{P}(X > x)\mathbb{P}(Y > y)$, or

equivalently, $\mathbb{P}(X \leq x, Y \leq y) \geq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$.

3. **Association.** X and Y are said to be positively associated if for every pair of functions a and b defined on \mathbb{R}^2 which are increasing in each of the arguments separately,

$$\text{cov}[a(X, Y), b(X, Y)] \geq 0.$$

Lai and Xie note that a direct verification of association is difficult [55]. It is often simpler to verify one or more of the conditions to follow, which are more stringent, and thus, imply association.

4. **Tail dependence.** Y is *right-tail increasing* in X , denoted as $\text{RTI}(Y|X)$, if $\mathbb{P}(Y > y|X > x)$ increases in x for all y . Similarly, Y is *left-tail decreasing* in X , written as $\text{LTD}(Y|X)$ if $\mathbb{P}(Y \leq y|X \leq x)$ decreases in x for all y .
5. **Stochastically increasing (SI).** Y is said to be stochastically increasing in x for all y , $\text{SI}(Y|X)$, if for every y , $\mathbb{P}(Y > y|X = x)$ is increasing in x . $\text{SI}(X|Y)$ can be defined in a similar manner. If Y is SI in X , $\mathbb{E}(Y|X = x)$ is also increasing in x .
6. **Total positivity of order 2.** Let X and Y have a joint density $f(x, y)$. Then f is said to be totally positive of order 2 (TP_2) if for all $x_1 < x_2, y_1 < y_2$,

$$f(x_1, y_1)f(x_2, y_2) \geq f(x_1, y_2)f(x_2, y_1)$$

TP_2 is also referred to as X and Y being likelihood ratio dependent (LRD).

Since these conditions were listed in the increasing order of stringency, $(6) \Rightarrow (5) \Rightarrow (4) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$. When the inequality signs of the relations described in (1) through (6) are reversed, we obtain analogous negative dependence concepts. Specifically, the duals of (2), (4), (5) and (6) are respectively called negative quadrant dependent, right tail decreasing/left

tail increasing dependence, stochastically decreasing dependence and reverse regular of order 2.

3.1.2 Measures of dependence

Measures of dependence quantify, in some particular manner, how closely the variables X and Y are related. Since a single number alone cannot completely explain the nature of dependence, a variety of measures are defined and used. The following list is not comprehensive, but represents some of the more important measures of dependence that have been proposed.

1. **Pearson's correlation.** This is a well studied measure in statistics and is presented here for completeness. Pearson's coefficient of correlation is given by,

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(x) \text{var}(Y)}}$$

It may be noted that ρ measures only the linear dependence. Furthermore, there exist well-known examples where X and Y are dependent, but $\rho = 0$. For example, Melnick and Tenenbein [62] have analyzed the following case. Let $X \sim \mathcal{N}(0, 1)$ and define Y such that for $\lambda > 0$

$$Y = \begin{cases} X & \text{if } |X| \leq \lambda \\ -X & \text{if } |X| > \lambda \end{cases} \quad (3.1)$$

We can verify that $Y \sim \mathcal{N}(0, 1)$, since

$$\mathbb{P}(Y \leq t) = \mathbb{P}(|X| \leq \lambda \wedge X \leq t) + \mathbb{P}(|X| > \lambda \wedge -X \leq t) \quad (3.2)$$

$$\begin{aligned} &= \mathbb{P}(|X| \leq \lambda \wedge X \leq t) + \mathbb{P}(|X| > \lambda \wedge X \leq t) \\ &= \mathbb{P}(X \leq t). \end{aligned} \quad (3.3)$$

where (3.2) follows from the symmetry of $\mathcal{N}(0, 1)$. Denote the p.d.f. of X as f_X and

CDF as F_X . The correlation coefficient can be calculated as

$$\begin{aligned}\rho = \mathbb{E}[XY] &= 2 \int_0^\lambda x^2 f_X(x) dx - 2 \int_\lambda^\infty x^2 f_X(x) dx \\ &= 4 \int_0^\lambda x^2 f_X(x) dx - 1\end{aligned}\tag{3.4}$$

Solving for λ by setting $\rho = 0$ in (3.4), Melnick and Tenenbein have obtained $\lambda \approx 1.54$; for this value of λ , in spite of X and Y being dependent, $\rho = 0$. Note that X and Y are not *jointly* normal, i.e., f_{XY} is not a bivariate normal p.d.f., and hence their dependence structure is not completely explained by ρ .

2. **Mutual information.** Mutual information between X, Y is defined as,

$$\mathcal{I}(X; Y) = \int_{\mathbb{R}^2} \log \left(\frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} \right) dF_{XY}(x, y),$$

and it measures the distance between the joint density and the product of marginals, i.e., the joint density if X, Y were independent. Multiinformation is the multivariate extension of mutual information proposed by Joe [40]. For the vector $\mathbf{X} \in \mathbb{R}^n, n > 2$,

$$\mathcal{I}(\mathbf{X}) = \int_{\mathbb{R}^n} \log \left(\frac{f_{\mathbf{X}}(\mathbf{x})}{\prod_i f_{X_i}(x_i)} \right) dF_{\mathbf{X}}(\mathbf{x}).$$

A normalization of the form $\delta^* = \sqrt{1 - \exp(-2\mathcal{I})}$ ensures that mutual information and multiinformation follow Rényi's postulates [77] for “an appropriate measure of dependence”. In particular, $\delta^* \in [0, 1]$.

3. **Rank correlations.** Rank correlations measure the dependence between rankings, rather than between actual values, of X and Y . Therefore, rank measures are unaffected by any increasing transformation of X and Y , while ρ is unaffected only by linear transformations. Kendall's tau (τ) and Spearman's rho (ρ_S) are widely used measures that fall in this category. For independent pairs of random variables (X_1, Y_1) and (X_2, Y_2) having the

same distribution as (X, Y) , concordance is defined as the condition that $(X_1 - X_2)(Y_1 - Y_2) \geq 0$ and discordance is defined as the condition that $(X_1 - X_2)(Y_1 - Y_2) < 0$. Kendall's tau is defined to be the difference between the probabilities of concordance and discordance:

$$\tau \triangleq \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) \geq 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

This definition is equivalent to,

$$\tau = \text{cov}[\text{sgn}(X_1 - X_2), \text{sgn}(Y_1 - Y_2)].$$

Kendall's tau is also a measure of total positivity: $\tau/2$ represents an average measure of the total positivity for f_{XY} , the joint density of X and Y .

Spearman's rho is defined as follows. Let $(X_i, Y_i), i = 1, 2, 3$ be three independent pairs of random variables with a common distribution function. Then,

$$\rho_S \triangleq 3 \{ \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) \geq 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0] \}.$$

Spearman's rho represents an average measure of quadrant dependence: $\rho_S \geq 0 \Rightarrow (X, Y)$ are PQD.

4. **Blomqvist's β .** This measure evaluates the dependence at the center of a distribution, where the center is defined by (\tilde{x}, \tilde{y}) , the medians of the two marginals. Hence, β is also referred to as the medial correlation coefficient. Blomqvist's β is defined as,

$$\beta = 2\mathbb{P}[(X - \tilde{x})(Y - \tilde{y}) > 0] - 1 \tag{3.5}$$

5. **Local measures of dependence.** Anscombe's quartet refers to four datasets that have

identical coefficients of correlation for four different sets of (X, Y) data pairs [5]. All four sets of data also have identical first and second order moments. While first constructed to demonstrate the importance of graphing data before analyzing it, the dataset also reveals the *global* nature of ρ , i.e., it is defined from the second moment, which is in turn an expectation evaluated over the entire plane. In other words, while global summary statistics are useful descriptors of the data, they often fall short of providing a complete picture about the true variability that exists in the data set. In fact, all of the above measures are global measures. Pairs (X, Y) and (X', Y') can have different distributions and yet have the same global measure. A local measure of dependence will allow one to compare the variation of dependence between the two pairs. Several local measures of dependence have been proposed in the literature, mostly as an extension of global dependence measures. Some of them are listed below.

- *Local correlation coefficient.* Let $\mu(x) = \mathbb{E}(Y|X = x)$, $\sigma^2(x) = \text{var}(Y|X = x)$ and $\beta(x) = \frac{\partial}{\partial x}\mu(x)$. The local correlation coefficient is then defined as

$$\rho(x) = \frac{\sigma_X \beta(x)}{[\sigma_X \beta(x)]^2 + \sigma^2(x)},$$

where σ_X is the standard deviation of X . When defined in this manner, $\rho(x)$ shares a few properties with its global counterpart: it takes values between 1 and -1, independence of X and Y implies that $\rho(x) = 0$ and $\rho(x) = \pm 1$ for almost all x is equivalent to Y being a function of X . It is also invariant to scaling, but is not marginal free. The latter point means that if we define $U = F_X(x)$ and $V = F_Y(y)$, the resulting $\rho(u)$ is different from $\rho(x)$.

- *Local τ and ρ_S .* Local measures of rank correlation exist, and are evaluated on an open neighborhood about a point of interest, (x_0, y_0) . The functional form is more easily defined using copulas, and is deferred to Section 3.2.

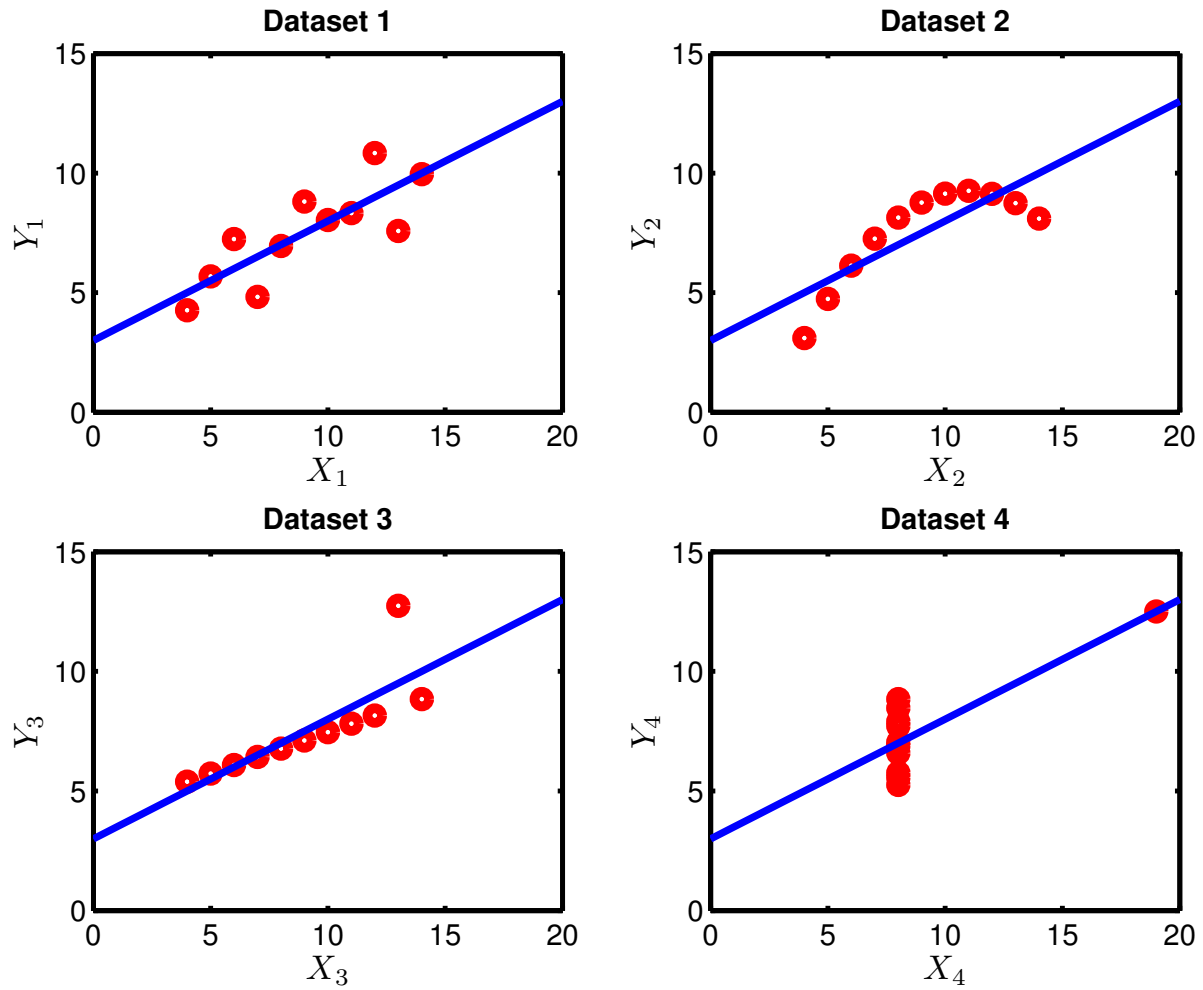


Fig. 3.1: Anscombe's quartet. All 4 datasets contain identical summary statistics: Mean of X_i , $\mu_{X_i} = 9$, variance of X_i , $\sigma_{X_i}^2 = 11$; mean of Y_i , $\mu_{Y_i} = 7.5$, variance of Y_i , $\sigma_{Y_i}^2 = 4.12$; correlation $\rho = 0.816 \forall i = 1, 2, 3, 4$.

- *Local measure of LRD.* An index that can be used to measure likelihood ratio dependence (LRD) locally is the second order partial derivative of the logarithm of the density function,

$$\gamma(x, y) = \frac{\partial^2}{\partial x \partial y} \log f_{XY}(x, y).$$

Recall that saying X and Y are LRD is synonymous with stating that $f_{XY}(x, y)$ is TP_2 . It can be shown that $\gamma(x, y) \geq 0 \forall x, y \Leftrightarrow f_{XY}(x, y)$ is TP_2 . This index has several attractive properties; significantly, $\gamma(x, y) = 0$ if and only if X and Y are independent. Furthermore, $\gamma(x, y)$ is marginal-free.

3.2 Copula theory

Copulas, typically defined as cumulative distribution functions (CDF), are parametric functionals that associate or “couple” disparate univariate marginal distributions to a multivariate distribution. The parametrization quantifies the dependence between the random variables over which the copula is defined. The dependence parameter is not explicitly specified in this section and is introduced in Section 4.2.2, as it is more relevant in the context of inference. Sklar’s theorem is an important result and specifies the framework necessary for copula-based inference [65]. Without loss of generality, the random variables are defined over $\bar{\mathbb{R}} \triangleq [-\infty, \infty]$.

Theorem 3.1 (Sklar’s Theorem). *A cumulative distribution function, $F_{\mathbf{Z}}$, is defined over the n -dimensional random vector $\mathbf{Z} = [Z_1, Z_2, \dots, Z_n]^T$ for which the corresponding marginal distribution functions are $F_{Z_1}, F_{Z_2}, \dots, F_{Z_n}$. There exists a copula C , such that for all $\mathbf{Z} \in \bar{\mathbb{R}}^n$,*

$$F_{\mathbf{Z}}(z_1, \dots, z_n) = C(F_{Z_1}(z_1), \dots, F_{Z_n}(z_n)) \quad (3.6)$$

If F_{Z_i} is continuous for $1 \leq i \leq n$, then C is unique, otherwise it is determined uniquely on $\text{Ran}F_{Z_1} \times \dots \times \text{Ran}F_{Z_n}$ where $\text{Ran}F_{Z_n}$ is the range of F_{Z_n} . Conversely, given a copula C and

univariate distributions F_{Z_1}, \dots, F_{Z_n} , F_Z as defined in (3.6) is a valid multivariate CDF with marginals F_{Z_1}, \dots, F_{Z_n} .

Note that (3.6) implies that the copula function is a joint distribution of uniformly distributed random variables. As a direct consequence of Sklar's Theorem, for continuous distributions, the joint p.d.f. is obtained by differentiating (3.6),

$$f_Z(\mathbf{z}) = \left\{ \prod_{i=1}^n f_{Z_i}(z_i) \right\} c(F_{Z_1}(z_1), \dots, F_{Z_n}(z_n)) \quad (3.7)$$

where $\mathbf{z} = [z_1, \dots, z_n]^\top$ and $c(\cdot)$, called the copula density, is obtained as the mixed derivative of C ,

$$c(\cdot) = \frac{\partial^n}{\partial u_1 \dots \partial u_n} C(u_1, \dots, u_n) \quad (3.8)$$

where, $u_i = F_{Z_i}(z_i) \sim \mathcal{U}(0, 1)$. Using (3.7), we can construct a joint density function with specified marginal densities.

Note that $C(\cdot)$ is a valid CDF and $c(\cdot)$ is a valid p.d.f. for uniformly distributed random variables, u_i . Many different types of signals have well-understood marginal sensor models, established either through physics-based theory or direct empirical evidence. An application specific understanding of dependence, however, is more difficult. Various families of copula functions, describing different types of dependence, have been proposed in the literature [65]. However, which copula function should be used for a given case is not very clear as different copula functions may characterize different types of dependence behavior among the random variables [60]. A brief summary of some popularly used copula functions is discussed next. In the following discussion, for notational brevity, we denote the n -tuple (u_1, \dots, u_n) as \mathbf{u} .

3.2.1 Summary of some copula functions

Copulas derived from distributions

Multivariate distribution functions specify dependence structures and copula functions can be derived from them. Two such copula functions are the Gaussian and the t copula functions that are derived from multivariate Gaussian and Student- t distributions respectively. Both specify dependence using the correlation matrix and are given as follows.

The Gaussian copula is defined as

$$C_{\mathcal{N}}(\mathbf{u}; \Sigma) = F_{\mathcal{N}}(F_{\mathcal{N}}^{-1}(u_1), \dots, F_{\mathcal{N}}^{-1}(u_n); \Sigma), \quad (3.9)$$

where, $F_{\mathcal{N}}(\cdot; \Sigma)$ denotes the multivariate normal CDF with correlation matrix Σ and $F_{\mathcal{N}}^{-1}$ denotes the inverse CDF of the standard normal. The corresponding copula density function is

$$c_{\mathcal{N}}(\mathbf{u}; \Sigma) = \frac{1}{\sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2} \boldsymbol{\omega}^{\top} (\Sigma - \mathbf{I}) \boldsymbol{\omega} \right\} \quad (3.10)$$

where $\boldsymbol{\omega} = [\omega_1, \dots, \omega_i, \dots, \omega_n]^{\top}$ with $\omega_i = F_{\mathcal{N}}^{-1}(u_i)$ and \mathbf{I} is the identity matrix.

Similarly, the t -copula is defined as

$$C_t(\mathbf{u}; \Sigma, \nu) = t_{\nu, \Sigma}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n)) = t_{\nu, \Sigma}(\xi_1, \dots, \xi_n) \quad (3.11)$$

where, $t_{\nu, \Sigma}$ is the multivariate Student- t distribution with correlation matrix Σ and ν degrees of freedom and t_{ν} denotes the univariate Student- t distribution with ν degrees of freedom. As $\nu \rightarrow \infty$, the t copula approaches the Gaussian copula. Let $\boldsymbol{\xi}$ denote the column vector of $\xi_i \forall i, 1 \leq i \leq n$. The density function for the t -copula is given by

$$c_t(\mathbf{u}; \Sigma, \nu) = \frac{\Gamma((\nu + n)/2) \Gamma(\nu/2)^{n-1} (1 + \nu^{-1} \boldsymbol{\xi}^{\top} \Sigma^{-1} \boldsymbol{\xi})^{-(\nu+n)/2}}{\sqrt{|\Sigma|} \Gamma((\nu + 1)/2)^n \prod_{i=1}^n (1 + \nu^{-1} \xi_i^2)^{-(\nu+1)/2}} \quad (3.12)$$

Both the Gaussian and the t copula functions belong to the elliptical family of copulas.

Archimedean copulas

Archimedean copulas, describing an n -variate CDF, are defined as follows,

$$C(\mathbf{u}; \phi) = \vartheta_{\phi}^{-1} \left(\sum_{i=1}^n \vartheta_{\phi}(u_i) \right) \quad (3.13)$$

where, $\vartheta : (0, 1] \mapsto [0, \infty)$ is a convex, strictly decreasing function with a positive second derivative with $\vartheta(1) = 0$. This function ϑ is referred to as the generator function and ϕ is the copula parameter specifying dependence. The inverse for the generator is defined as,

$$\vartheta^{-1}(s) = \begin{cases} \vartheta^{-1}(s) & \text{for } 0 \leq s \leq \vartheta(0) \\ 0 & \text{for } \vartheta(0) < s < \infty \end{cases} \quad (3.14)$$

While for statistical inference, the copula density is more useful, it is more difficult to derive a usable expression for every Archimedean copula. Using (3.8), we can write

$$c(\mathbf{u}; \phi) = (\vartheta_{\phi}^{-1})^{(n)} \left(\sum_{i=1}^n \vartheta_{\phi}(u_i) \right) \prod_{i=1}^n \frac{\partial}{\partial u_i} \vartheta_{\phi}(u_i) \quad (3.15)$$

where the superscript (n) refers to the n -th order partial derivative over $\vartheta_{\phi}(u_i)$. For a bivariate Archimedean copula this resolves to

$$c(u_1, u_2) = - \frac{\vartheta_{\phi}''(C(u_1, u_2)) \vartheta_{\phi}'(u_1) \vartheta_{\phi}'(u_2)}{[\vartheta_{\phi}'(C(u_1, u_2))]^3} \quad (3.16)$$

The Clayton, Frank and Gumbel copulas are commonly used examples of the Archimedean copula family and are defined next.

Clayton copula

The generator function for the Clayton copula is

$$\vartheta_\phi(u) = \frac{1}{\phi} (u^{-\phi} - 1) \quad \phi \in [-1, \infty) \setminus \{0\} \quad (3.17)$$

and, therefore, the copula CDF is given by

$$C_{\text{Cl}}(\mathbf{u}; \phi) = \left(\sum_{i=1}^n u_i^{-\phi} - n + 1 \right)^{-\frac{1}{\phi}}, \quad \phi \in [-1, \infty) \setminus \{0\} \quad (3.18)$$

and the copula density function can be obtained upon differentiation as

$$c_{\text{Cl}}(\mathbf{u}; \phi) = \phi^n \frac{\Gamma\left(\frac{1}{\phi} + n\right)}{\Gamma\left(\frac{1}{\phi}\right)} \left(\prod_{i=1}^n u_i^{-\phi-1} \right) \left(\sum_{i=1}^n u_i^{-\phi} - n + 1 \right)^{-\frac{1}{\phi}-n}. \quad (3.19)$$

Frank copula

The Frank copula uses the generator function

$$\vartheta_\phi(u) = -\log \frac{\exp\{-\phi u\} - 1}{\exp\{-\phi\} - 1}, \quad \phi \in \mathbb{R} \setminus \{0\} \quad (3.20)$$

which leads to the associated copula CDF

$$C_{\text{Fr}}(\mathbf{u}; \phi) = -\frac{1}{\phi} \log \left(1 + \frac{\prod_{i=1}^n [\exp\{-\phi u_i\} - 1]}{\exp\{-\phi\} - 1} \right), \quad \phi \in \mathbb{R} \setminus \{0\}. \quad (3.21)$$

The n -variate copula density is difficult to derive. Archimedean copulas are more useful in their bivariate form, and as we will see in Chapter 5, construction of multivariate copulas, using bivariate elements, leads to a better model, in general. The bivariate copula density is

given by setting $n = 2$ and twice differentiating the copula CDF in (3.21). Therefore,

$$c_{\text{Fr}}(u_1, u_2; \phi) = \frac{\phi(1 - \exp\{-\phi\}) \exp\{-\phi(u_1 + u_2)\}}{[1 - \exp\{-\phi\} - (1 - \exp\{-\phi u_1\})(1 - \exp\{-\phi u_2\})]^2}. \quad (3.22)$$

Gumbel copula

The function

$$\vartheta_\phi(u) = (-\log u)^\phi, \quad \phi \in [1, \infty) \quad (3.23)$$

generates the Gumbel copula CDF

$$C_{\text{Gu}}(\mathbf{u}; \phi) = \exp \left\{ - \left(\sum_{i=1}^n (-\ln u_i)^\phi \right)^{\frac{1}{\phi}} \right\}. \quad (3.24)$$

The for $n = 2$, we obtain the corresponding bivariate copula density function

$$\begin{aligned} c_{\text{Gu}}(u_1, u_2; \phi) &= \frac{C(u_1, u_2; \phi)}{u_1 u_2} [(-\log u_1)^\phi + (-\log u_2)^\phi]^{-2(1-\frac{1}{\phi})} [(\log u_1)(\log u_2)]^{\phi-1} \\ &\quad \times \{1 + (\phi - 1)[(-\log u_1)^\phi + (-\log u_2)^\phi]^{-\frac{1}{\phi}}\} \end{aligned} \quad (3.25)$$

In addition to these copulas, we also note that independence is also a valid Archimedean copula, with $-\log u$ as the generator function.

3.2.2 Copulas and measures of dependence

For a joint bivariate CDF expressed as a copula, some interesting observations can be made about the various measures of dependence introduced in Section 3.1.2. For the random pair

(X, Y) , Kendall's τ and Spearman's ρ are respectively expressed as the following expectations:

$$\tau = 4 \mathbb{E}[C(F_X(x), F_Y(y))] - 1 \quad (3.26)$$

$$\rho_S = 12 \mathbb{E}[F_X(x)F_Y(y)] - 3 \quad (3.27)$$

For the case of elliptical copulas, parametrized by the matrix $\Sigma = [\rho_{\Sigma}(i, j)]$,

$$\rho_{\Sigma}(i, j) = \sin\left(\frac{\pi\tau_{ij}}{2}\right), \quad (3.28)$$

where $\tau_{i,j}$ is the Kendall's τ evaluated for the pair (U_i, U_j) .

Blomqvist's β defined in Eq. (3.5) can be expressed in terms of the bivariate copula, C , for the pair (X, Y) as,

$$\beta = 4F_{XY}(\tilde{x}, \tilde{y}) - 1 = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \quad (3.29)$$

Nelsen [65] notes that although β depends only on the value of the copula at the center of $[0, 1] \times [0, 1]$, it can provide good approximations of τ and ρ_S using, e.g., a Maclaurin series expansion.

Local measures of dependence discussed earlier also reveal interesting properties when expressed in terms of a copula. When the expectation is restricted to an open neighborhood $V(x_0, y_0)$ local forms of τ and ρ_S are defined as,

$$\tau(x_0, y_0) = 4 \iint_{V(x_0, y_0)} C(u, v) du dv - 1 \quad (3.30)$$

$$\rho_S(x_0, y_0) = 12 \iint_{V(x_0, y_0)} (C(u, v) - uv) du dv \quad (3.31)$$

In Section 3.1.2, a local measure of likelihood ratio dependence (LRD) was defined as

$$\gamma(x, y) = \frac{\partial^2}{\partial x \partial y} f_{XY}(x, y)$$

and it was noted that this measure is marginal free. Consequently, $\gamma(x, y)$ equals $\gamma(u, v)$, where

$$\gamma(u, v) = \frac{\partial^2}{\partial u \partial v} c(u, v), \quad F_X(x) = u, F_Y(y) = v, \quad (3.32)$$

and $c(u, v)$ is the copula density function for copula C .

3.2.3 Tail dependence coefficients as a measure of extremal dependence

Extremal dependence is the characterization of statistical co-movement for extreme values of multivariate data. In the context of bivariate data, tail dependence coefficients are a natural measure of extremal dependence. Two measures, the upper and lower tail dependence coefficients, have been defined in the literature, and they measure the amount of dependence in the upper and lower quadrant tails of the support of the random vector. Let $[X, Y]$ be a vector of continuous random variables with marginal CDFs F and G . Let $C(F(X), G(Y))$ be a bivariate copula distribution function. Then,

$$\lambda_U \triangleq \lim_{u \nearrow 1} \mathbb{P}(Y > G^{-1}(u) | X > F^{-1}(u)) \quad (3.33)$$

$$= \lim_{u \nearrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \quad (3.34)$$

$$\lambda_L \triangleq \lim_{u \searrow 0} \mathbb{P}(Y \leq G^{-1}(u) | X \leq F^{-1}(u)) \quad (3.35)$$

$$= \lim_{u \searrow 0} \frac{C(u, u)}{u} \quad (3.36)$$

Using these relations, one can show that, for the Gaussian copula

$$\lambda_L = \lambda_U = 2 \lim_{x \rightarrow -\infty} F_{\mathcal{N}} \left(x \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) = 0$$

The t -copula on the other hand exhibits non-zero upper and lower tail dependence, i.e.,

$$\lambda_L = \lambda_U = 2t_{\nu+1} \left(-\sqrt{\nu+1} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right)$$

where $t_{\nu+1}$ denotes the CDF of a univariate t distribution with $\nu+1$ degrees of freedom. Hence, for large values of ρ and small values of ν the t -copula exhibits strong tail dependence.

3.3 Summary

In this chapter, we have seen that copulas are able to provide a complete characterization of statistical dependence, largely because of their functional nature. Additionally, for many families, there exists a one-to-one relationship between the copula dependence parameters and nonparametric rank-based measures of dependence, such as Kendall's tau. The use of these measures in inference leads to a large savings in computational effort, as compared to optimal approaches such as maximum likelihood. The copula-based approach also allows us to characterize extremal dependence through the concept of tail-dependence. Selecting and using copulas that possess non-zero tail-dependence plays an important role in inference problems. These issues are discussed in further detail in the next chapter, in the context of inference using heavy-tailed α -stable sensor models.

CHAPTER 4

DETECTION OF DEPENDENT HEAVY-TAILED DATA

In this chapter, we take the first steps at formulating and deriving the theory for spatially dependent heavy-tailed signals, using a copula-based approach for dependence modeling. When extreme value measurements occur at a significantly greater frequency than is attributable to distributions that decay exponentially in the tail, often polynomial tail-decay models provide an appropriate fit. These models can accommodate the typical “spiky” signatures in the signal measurements and such data are often said to be *fat-tailed* or *heavy-tailed*. Examples of such data are seen in applications such as climatology [24], finance [99], and well-established signal processing applications such as radar, communications and image processing [3, 13, 45]. The co-occurrence of such (rare) extreme-valued data is sometimes symptomatic of a catastrophic event, and its detection, therefore, needs appropriate modeling tools.

The heavy-tailed characteristics in these applications are often modeled using a class of functions known as α -stable distributions. Excluding the Lévy, Cauchy and Gaussian distributions, the α -stable family does not admit a closed-form probability density function (p.d.f.). They are instead defined using characteristic functions [67, 81]. This chapter examines the

problem of detection of spatially dependent α -stable signals. We consider a setup where the data coming from all sensors are α -stable distributed, but are non-identically distributed. In this sense, the sensors are heterogeneous. As discussed in Chapter 1, the cause for this heterogeneity could be multifarious.

4.1 Introduction

The α -stable model is motivated by the empirical observation that several non-Gaussian phenomena exhibit a power-law decay model with a tail of the type $|x|^{-\alpha-1}$, $\alpha \in (0, 2)$; α is referred to as the tail-index. Further, Gnedenko and Kolmogorov [32] proved a *generalized* central limit theorem (CLT) for random variables that possess this power-law tail decay property. This theorem states that the limiting distribution of the sum of power-law heavy-tailed distributed random variables tends to the class of α -stable distributions. In addition to α , this class of distributions has three additional parameters corresponding to location (δ), scale (γ) and skewness (β). This allows for flexible modeling of various types of non-Gaussian data. If $\beta = 0$, one obtains an important special case called the symmetric α -stable distribution, often denoted as $S\alpha S$. A formal definition and brief introduction to the theory of α -stable distributions is presented in Section 4.2.1.

Introductory discussions, from a signal processing perspective, on α -stable processes have focused on independent and identically distributed (IID) formulations [6, 84]. Detection in the presence of IID $S\alpha S$ noise was investigated using Bayesian and Neyman-Pearson approaches [95, 96], where fractional lower order moments (FLOM) were used to estimate unknown parameters. Kuruoğlu et al. [54] have used a mixture of Gaussian approximation for $S\alpha S$ noise. Swami and Sadler [91] used higher order statistics for estimating and detecting signals in $S\alpha S$ noise with unknown parameters. More recently, different authors have explored the use of α -stable models in distributed detection [73], acoustic tracking [110], anomaly detection [85], wireless communications [76, 80] and biomedical applications [56]. An ex-

tensive bibliography on α -stable distributions/processes and its applications is maintained by Nolan [68].

In this chapter, we consider a detection problem using data from sensors configured in a parallel topology. As indicated in Chapter 1, heterogeneous sensors observe a common phenomenon. Their observations may be made over an arbitrary domain of measurement. For example, these measurements may represent a time series (temporal measurements), a sequence of spectral coefficients (measurements in frequency domain), or some other feature vector. In their respective measurement domains, sensor observations are modeled as IID α -stable random variables (e.g., temporally independent or independent spectral coefficients). The sensor signal model is kept quite general, i.e., we do not explicitly specify whether the phenomenon of interest is embedded in IID α -stable noise or if the α -stable model characterizes the dynamics of the phenomenon itself.

Since the sensors jointly measure the same process, their measurements are spatially dependent (i.e., across sensors). We use copulas to model this dependence (see Chapter 3). This α -stable-copula model serves as the focal point of our investigation of detection of dependent heavy-tailed data. The generality of our signal model and the copula-based dependence formulation distinguishes this work from previous works, such as [73], which have specifically considered conditionally independent sensor observations embedded in α -stable noise.

Multivariate α -stable models have also been defined and used for inference on random vectors with heavy tails (e.g., see [69, 70, 75, 81]). A multivariate α -stable model generalizes the univariate α -stable law and is defined using a joint characteristic function. Consequently, with the exception of a few applications¹, obtaining the resultant α -stable marginal densities is not computationally tractable. In contrast, the copula approach allows for the *synthesis* of a joint distribution based on pre-specified, possibly heterogeneous, marginal models. Recall that copulas are parametric probability distributions that couple univariate marginals to gener-

¹see Nolan [68, 69] and references cited therein

ate a valid joint distribution that incorporates statistical dependence. In this chapter, we utilize several families of copula functions, which can characterize non-linear and asymmetric dependencies. Copula-based methods of inference also scale well across multisensor or multidimensional formulations. This should be contrasted with completely nonparametric formulations, such as learning-based techniques, which are known to suffer from scalability issues stemming from the curse of dimensionality.

In the following sections, we develop the idea of distributed signal detection, using a copula based characterization of dependence, for α -stable data. Section 4.2 lays out the canonical signal model for spatially dependent α -stable data. The detection problem is formulated in Section 4.3, and variations of the likelihood ratio test under the Neyman-Pearson framework are studied. The proposed detection schemes are applied to simulated data and the results thus obtained are discussed in Section 4.4.

4.2 Signal Model

We consider a two-sensor system, where each sensor transmits its analog measurements or observation data to the fusion center (FC). The two-sensor restriction is without loss of generality: the theory developed in this chapter (Propositions 4.1, 4.3 and 4.4) readily extends to multiple sensors and, significantly, our main conclusions do not depend upon the number of sensors. The two-sensor formulation allows us to minimize the notational complexity in the exposition of the theory.

Sensor $i \in \{1, 2\}$ transmits $\{x_{ij}\}_{j=1}^N$, a sequence of N IID measurements, to the FC. Each x_{ij} is a realization of the random variable X_i , where j indexes the measurement domain, which can be time, frequency, or any other feature. The sensor model is the p.d.f. f_{X_i} , which is characterized using α -stable distributions (Section 4.2.1).

Denote the j -th observation pair as $\mathbf{x}_j = [x_{1j}, x_{2j}]^\top$ where $[\cdot]^\top$ denotes matrix/vector transpose. In general, the random vector $\mathbf{X} = [X_1, X_2]^\top$, has a joint density $f_{\mathbf{X}}(\mathbf{x}_j) \neq$

$f_{X_1}(x_{1j}) \cdot f_{X_2}(x_{2j})$, i.e., sensor observations are spatially dependent. This inter-sensor dependence is modeled using copulas (Section 3.2). For $\mathbf{x} = \{\mathbf{x}_j\}_{j=1}^N$, $f_{\mathbf{X}}(\mathbf{x}) = \prod_j f_{\mathbf{X}}(\mathbf{x}_j)$, i.e., sensor data are independent across j . It is not necessary that $f_{X_1} = f_{X_2}$, i.e., the sensors are *heterogeneous*. The α -stable p.d.f. f_{X_i} is also referred to as the *marginal density* since we can obtain each sensor model by marginalizing $f_{\mathbf{X}}$; α -stable parameters corresponding to marginal p.d.f.s are called the *marginal parameters*.

The FC uses the received data to calculate a test-statistic, which is compared to a threshold. Under the Neyman-Pearson framework, the threshold is chosen such that the probability of detection, P_D , is maximized under a constraint on P_F , the probability of false alarm.

4.2.1 Stable distributions

We model X_i as an α -stable random variable. An α -stable distribution (also referred to simply as a *stable distribution*), does not necessarily have a closed-form p.d.f. They are defined in closed-form by their characteristic function (CF),

$$\varphi_{X_i}(t) = \exp(-\gamma^\alpha |t|^\alpha B_\alpha(t) + \mathbf{i}\delta t) \quad (4.1)$$

$$B_\alpha(t) = \begin{cases} [1 - \mathbf{i}\beta \tan(\pi\alpha/2) \operatorname{sgn}(t)], & \text{for } \alpha \neq 1 \\ [1 + \mathbf{i}\beta(2/\pi) \operatorname{sgn}(t) \log |t|], & \text{for } \alpha = 1 \end{cases} \quad (4.2)$$

where $\mathbf{i} = \sqrt{-1}$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ and $\delta \in \mathbb{R}$. The parameters α , β , γ and δ are, respectively, the tail-index, location, dispersion and skewness parameters. The CF, $\varphi_{X_i}(t)$, and p.d.f., $f_{X_i}(x_{ij})$, are Fourier transform pairs.

We denote the distribution of X_i as

$$X_i \sim \mathcal{S}(\alpha, \beta, \gamma, \delta). \quad (4.3)$$

The *standard form* refers to the case where $\delta = 0$ and $\gamma = 1$ so that, $\gamma^{-1/\alpha}(X_i - \delta) \sim \mathcal{S}(\alpha, \beta, 1, 0)$. The support for f_{X_i} depends on the values of α, β and δ [67],

$$\text{supp}[f_{X_i}] = \begin{cases} [\delta, \infty) & \text{for } \alpha < 1, \beta = 1, \\ (-\infty, \delta] & \text{for } \alpha < 1, \beta = -1, \\ \mathbb{R} & \text{otherwise.} \end{cases} \quad (4.4)$$

Remarks

Some special cases and properties of α -stable distributions are as follows:

- *Closed-form p.d.f.* A closed-form p.d.f. exists for three special cases: Cauchy ($\alpha = 1, \beta = 0$), Lévy ($\alpha = 0.5, \beta = 1$), and normal ($\alpha = 2$) distributions.
- *Existence of moments.* The m -th order moment exists only if $m \in (0, \alpha)$. For example, for the Cauchy distribution neither mean nor variance is defined since $\alpha = 1$.
- *Fractional order moments.* Analogous to L_p norms for non-integer values of $p \leq 2$, typically considered in robust control, the p -th order fractional moments (see [84]) of an α -stable random variable can be defined as,

$$\mathbb{E}[|X_{ij}|^p], \quad \text{for } p < \alpha.$$

For $p \geq \alpha$, $\mathbb{E}[|X_{ij}|^p] = \infty$. These p -th order moments are also called fractional lower order moments (FLOM). Parameter estimation based on FLOM has been an active area of research [6].

- *Symmetry.* As noted in Section 4.1, $\beta = \delta = d$ implies a symmetric distribution about the median d . This is also called a symmetric α -stable distribution, denoted as $S\alpha S$. For the $S\alpha S$ case, when the mean is not admissible, δ corresponds to the median [95].

4.2.2 Dependent stable signals

For the two sensor problem under consideration, recall that X_i is distributed as in (4.3), i.e., $X_1 \sim \mathcal{S}(\alpha_1, \beta_1, \gamma_1, \delta_1)$ and $X_2 \sim \mathcal{S}(\alpha_2, \beta_2, \gamma_2, \delta_2)$. Using the vector notation

$$\boldsymbol{\psi}_i = [\alpha_i, \beta_i, \gamma_i, \delta_i]^\top, \quad i = 1, 2, \quad (4.5)$$

the marginal density for the j -th observation from sensor i is $f_{X_i}(x_{ij}; \boldsymbol{\psi}_i)$. Consider an arbitrary, possibly unknown copula, c , parametrized by a d -dimensional column vector, ϕ_c . The dimension, d , and the properties of ϕ_c depend on the definition of the specific copula, c . Denote the probability integral transform for the copula argument as,

$$u_{ij}(\boldsymbol{\psi}_i) \triangleq F_{X_i}(x_{ij}; \boldsymbol{\psi}_i), \quad i = 1, 2 \quad (4.6)$$

Thus, for dependent α -stable signals, (3.7) can be rewritten as,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{j=1}^N f_{X_1}(x_{1j}; \boldsymbol{\psi}_1) f_{X_2}(x_{2j}; \boldsymbol{\psi}_2) \\ &\quad \times c(F_{X_1}(x_{1j}; \boldsymbol{\psi}_1), F_{X_2}(x_{2j}; \boldsymbol{\psi}_2); \phi_c) \\ &= \prod_{j=1}^N f_{X_1}(x_{1j}; \boldsymbol{\psi}_1) f_{X_2}(x_{2j}; \boldsymbol{\psi}_2) \\ &\quad \times c(u_{1j}(\boldsymbol{\psi}_1), u_{2j}(\boldsymbol{\psi}_2); \phi_c) \end{aligned} \quad (4.7)$$

where the column vector

$$\boldsymbol{\theta} = [\boldsymbol{\psi}_1 \ \boldsymbol{\psi}_2 \ \phi_c]^\top$$

is contained in the parameter space, Θ_c , defined as the product set of respective component marginal and copula parameter spaces, Ψ_i and Φ_c . That is,

$$\Theta_c \triangleq \Psi_1 \times \Psi_2 \times \Phi_c$$

This serves as the canonical signal model or data generating process (DGP), which is further qualified in terms of the null and alternative hypotheses for the detection problem. All notations leading to the DGP, as well as those appearing in Section 4.3, are summarized in Table 4.1.

4.3 The detection problem

We formulate the detection problem as a test of hypotheses

$$H_0: f_{\mathbf{X}}^0 = f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}_0) \text{ vs. } H_1: f_{\mathbf{X}}^1 = f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}_1), \quad (4.8)$$

$f_{\mathbf{X}}^0 \neq f_{\mathbf{X}}^1, \forall j = 1, 2, \dots, N$. The parameters under the null and alternative hypotheses are, respectively,

$$\boldsymbol{\theta}_0 = \begin{bmatrix} \boldsymbol{\psi}_{10} & \boldsymbol{\psi}_{20} & \phi_{c_0} \end{bmatrix}^T \quad (4.9)$$

and,

$$\boldsymbol{\theta}_1 = \begin{bmatrix} \boldsymbol{\psi}_{11} & \boldsymbol{\psi}_{21} & \phi_{c_1} \end{bmatrix}^T. \quad (4.10)$$

In (4.8), we assume that H_0 is completely specified, whereas H_1 is composite. Such a formulation is frequently encountered in applications such as anomaly detection, where the “normal” operational state of a process is known *a priori*. Specifically, $\boldsymbol{\theta}_0$ is a fixed and known point in the parameter space Θ_{c_0} , defined for the (known) copula c_0 . $\boldsymbol{\theta}_1 \in \Theta_{c_1}$ is deterministic but unknown such that the distribution parameters as well as the copula function under H_1 , c_1 , are unknown. The space Θ_{c_1} is not defined completely since c_1 is assumed to be unknown. Therefore, the formulation in (4.8) leads to a test over the parameter space, as well as the space of copula functions.

In order to simplify the problem, we consider a copula library, \mathcal{C} , containing candidate copulas, defined over an indexing set, \mathcal{M} . For the applications discussed in Section 4.4.3, we use the copulas listed in Table 4.2. These are among the most commonly applied copulas in the

Table 4.1: Symbols and Notations

<i>Notation</i>	<i>Description</i>
j	Measurement index, $j = 1, 2, \dots, N$
x_{ij}	j -th observation from i -th sensor, $i = 1, 2$
X_i	α -stable random variable corresponding to x_{ij}
ψ_i or ψ_{ik}	Vector of α -stable parameters for sensor i and hypothesis $k \in \{0, 1\}$, where specified; see (4.5).
$f_{X_i}(x_{ij}; \psi_i)$	Sensor model or p.d.f. of X_i
$F_{X_i}(x_{ij}; \psi_i)$	CDF of X_{ij} ; also see (4.6)
$\mathbf{X}, \mathbf{x}_j, \mathbf{x}$	Random vector $[X_1, X_2]^\top$, its j -th sample realization, $\mathbf{x}_j = [x_{1j}, x_{2j}]^\top$, and the sequence, $\mathbf{x} = \{\mathbf{x}_j\}_{j=1}^N$.
ϕ_c	Dependence parameter for copula c
$\boldsymbol{\theta}$ or $\boldsymbol{\theta}_k$	Joint parameter vector $[\psi_1, \psi_2, \phi_c]^\top$ or, specifically, $[\psi_{1k}, \psi_{2k}, \phi_{c_k}]^\top$ under H_k
$f_{\mathbf{X}}$	Joint p.d.f. of \mathbf{X} expressed as a product of α -stable marginals f_{X_1} , f_{X_2} and copula c ; see (4.7).
$f_{\mathbf{X}}^k$	Joint p.d.f. under hypothesis k .
\mathcal{C}	Copula function library; see Table 4.2
c^*	Copula selected using ML-based copula selection
$\hat{\psi}_i, \hat{\phi}_{c^*}$	ML estimates of marginal parameters ψ_i and copula parameter ϕ_{c^*}
$\hat{\boldsymbol{\theta}}_*$	Concatenated vector $[\hat{\psi}_1, \hat{\psi}_2, \hat{\phi}_{c^*}]^\top$
$f_{\mathbf{X}}^*$	Joint p.d.f. obtained using c^* as the copula with ML parameter estimates, i.e., $f_{\mathbf{X}}(\mathbf{x}_j; \hat{\boldsymbol{\theta}}_*)$
$\tilde{\boldsymbol{\theta}}_k$	Pseudo-true value, evaluated under hypothesis H_k , when $f_{\mathbf{X}}^k$ is not the data generating process
$\tilde{f}_{\mathbf{X}}^k$	Joint p.d.f. evaluated at $\tilde{\boldsymbol{\theta}}_k$, i.e., $f_{\mathbf{X}}(\mathbf{x}_j; \tilde{\boldsymbol{\theta}}_k)$

Table 4.2: Library of copula functions

<i>Copulas</i>	<i>Parametric CDF</i>	<i>Parameter range</i>
Gaussian	$F_G(F_G^{-1}(x_1), \dots, F_G^{-1}(x_m); \Sigma),$	$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$ $\rho \in [-1, 1]$
Student- t	$t_{\nu, \Sigma}(t_{\nu}^{-1}(x_1), \dots, t_{\nu}^{-1}(x_m)),$	$\nu \geq 3$
Clayton	$\left(\sum_{i=1}^m u_i^{-\phi} - 1 \right)^{-1/\phi}$	$\phi \in [-1, \infty) \setminus \{0\}$
Frank	$-\frac{1}{\phi} \log \left[\frac{1 + (\prod_{i=1}^m e^{-\phi u_i} - 1)}{(e^{-\phi} - 1)} \right]$	$\phi \in \mathbb{R} \setminus \{0\}$
Gumbel	$\exp \left\{ - \left(\sum_{i=1}^m (-\ln u_i)^{\phi} \right)^{\frac{1}{\phi}} \right\}$	$\phi \in [1, \infty)$
Product	$\prod_{i=1}^m u_i$	—

$F_G(\mathbf{x}; \Sigma)$: multivariate normal CDF with mean $\mathbf{0}$ and covariance Σ ;

$F_G^{-1}(x_i)$: inverse univariate normal CDF with mean 0 and variance 1;

$t_{\nu, \Sigma}$: multivariate Student- t CDF; t_{ν}^{-1} : inverse CDF of univariate Student- t

literature [17]. Note that each copula is defined as a bivariate CDF; the corresponding density function is obtained by using (3.8).

Since the copula corresponding to the DGP under H_1 is unknown, a best fit is selected from the functions contained in \mathcal{C} . Therefore, the hypothesis testing problem implicitly contains a model selection component in the formulation, in which we attempt to identify the “best” copula, $c^*(\cdot; \phi_{c^*}^*)$, for the alternative hypothesis, where $\{c^*(\cdot; \phi_{c^*}^*) | \phi_{c^*}^* \in \Phi_{c^*}\} \in \mathcal{C}$. In developing the theory, we assume that the “true” copula is contained in \mathcal{C} , i.e., our models are *well specified*. The effect of model misspecification in the context of copula-based hypothesis testing has been addressed by Iyengar [36]. In general, the selected copula, c^* , may not admit any parameter which can also describe the copula model under the null hypothesis, $c_0(\cdot; \phi_{c_0})$. That is, there may not exist $\phi_{c^*}^* \in \Phi_{c^*}$ such that $c^*(\cdot; \phi_{c^*}^*) = c_0(\cdot; \phi_{c_0})$. Therefore, our formulation also considers the more general case of testing *non-nested hypotheses* [20, 21, 74, 100, 105].

The hypothesis testing problem is solved under the Neyman-Pearson framework, i.e., we

seek to design tests that operate under a false-alarm constraint. We use the generalized likelihood ratio test (GLRT) as the starting point and investigate its properties. The GLR test-statistic is modified, to accommodate uncertainty about c_1 , by also maximizing over \mathcal{C} , so that

$$T_{\text{GLR}} = \log f_{\mathbf{X}}(\mathbf{x}; \hat{\boldsymbol{\theta}}_*) - \log f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}_0) \quad (4.11)$$

$$\begin{aligned} &= \left[\sum_{j=1}^N \log f_{X_1}(x_{1j}; \hat{\boldsymbol{\psi}}_1) - \log f_{X_1}(x_{1j}; \boldsymbol{\psi}_{10}) \right] \\ &\quad + \left[\sum_{j=1}^N \log f_{X_2}(x_{2j}; \hat{\boldsymbol{\psi}}_2) - \log f_{X_2}(x_{2j}; \boldsymbol{\psi}_{20}) \right] \\ &\quad + \left[\sum_{j=1}^N \log c^*(u_{1j}(\hat{\boldsymbol{\psi}}_1), u_{2j}(\hat{\boldsymbol{\psi}}_2); \hat{\phi}_{c^*}) \right. \\ &\quad \left. - \log c_0(u_{1j}(\boldsymbol{\psi}_{10}), u_{2j}(\boldsymbol{\psi}_{20}); \phi_{c_0}) \right] \end{aligned} \quad (4.12)$$

$$\text{and, } \hat{\boldsymbol{\psi}}_i = \arg \sup_{\boldsymbol{\psi}_i \in \Psi_i} \sum_{j=1}^N \log f_{X_i}(x_{ij}; \boldsymbol{\psi}_i), \quad i = 1, 2. \quad (4.13)$$

Given an arbitrary indexing set, \mathcal{M} , for the copula library, \mathcal{C} , $c^* \equiv c_{m^*}$ such that for any $c_m \in \mathcal{C}, m \in \mathcal{M}$

$$m^* = \arg \max_{m \in \mathcal{M}} \left\{ \ell_{c_m}(\hat{\phi}_m) \mid \hat{\phi}_m = \arg \sup_{\phi_m \in \Phi_m} \ell_{c_m}(\phi_m) \right\}, \quad (4.14)$$

$$\ell_{c_m}(\phi_m) = \sum_{j=1}^N \log c_m(u_{1j}(\hat{\boldsymbol{\psi}}_1), u_{2j}(\hat{\boldsymbol{\psi}}_2); \phi_m). \quad (4.15)$$

In (4.11), $\hat{\boldsymbol{\theta}}_*$ is obtained by estimating the marginal parameters, $\boldsymbol{\psi}_i$, independently, prior to obtaining $\hat{\phi}_m$, as in (4.14) and (4.15). This two-step procedure is known as the *inference for margins* (IFM) method [17] and is different from estimating the marginal and copula parameters simultaneously. It follows from (4.14) that $\hat{\phi}_{c^*} \in \Phi_{c^*}$, in (4.12), is the maximum likelihood

estimate (MLE) of c^* . The decision rule is

$$T_{\text{GLR}} \underset{H_0}{\overset{H_1}{\geq}} \eta_a, \quad (4.16)$$

where η_a is the threshold that satisfies the constraint $P_F = a$. Deriving the distribution for T_{GLR} for finite N is very difficult. Asymptotic distributions, however, may be derived. In the analysis that follows, we denote $f_{\mathbf{X}}(\mathbf{x}_j; \hat{\boldsymbol{\theta}}_*)$ as $f_{\mathbf{X}}^*$ and include the subscript N in the notation for finite-sample statistics to emphasize dependence on sample size, as necessary.

4.3.1 Nested hypotheses or nested copula models

In general, for arbitrary hypotheses H and K , H is said to be nested in K if it is possible to derive H from K “either by means of an exact set of parametric restrictions or as a result of a limiting process” [74]. To define nesting in a more precise manner, it is helpful to define a *model* as a set of p.d.f.s indexed over admissible parameter values. The p.d.f. form of a nested model [100] is stated in Definition 4.1. Based on this, we formally define a *nested copula model*, which allows us to derive asymptotic results for our formulation.

Definition 4.1 (Nested model). For a continuous random vector $\mathbf{Z} \in \mathbb{Z} \subset \mathbb{R}^n$, given two models, $\mathcal{F}_0 \triangleq \{f_0(\mathbf{z}; \boldsymbol{\theta}_0) \mid \boldsymbol{\theta}_0 \in \Theta_0\}$ and $\mathcal{F}_1 \triangleq \{f_1(\mathbf{z}; \boldsymbol{\theta}_1) \mid \boldsymbol{\theta}_1 \in \Theta_1\}$, where f_0 and f_1 are arbitrary p.d.f.s of \mathbf{Z} , \mathcal{F}_0 is said to be *nested* in \mathcal{F}_1 if and only if $\mathcal{F}_0 \subset \mathcal{F}_1 \forall \mathbf{z} \in \mathbb{Z}$.

Definition 4.2 (Nested copula model). A copula family or model $\mathcal{C}_0 \triangleq \{c_0(u, v; \phi_0) \mid \phi_0 \in \Phi_0\}$ is nested in copula model $\mathcal{C}_1 \triangleq \{c_1(u, v; \phi_1) \mid \phi_1 \in \Phi_1\}$ if and only if $\mathcal{C}_0 \subset \mathcal{C}_1$ for $(u, v) \in [0, 1]^2$ almost everywhere.

Nested copulas and nested models are related to each other through the following lemma.

Lemma 4.1. For $k = 0, 1$, arbitrary continuous p.d.f.s f_k , g_k , h_k and copulas c_k , define the

models

$$\mathcal{F}_k = \{f_k(x; \psi_{f_k}) \mid \psi_{f_k} \in \Psi_{f_k}\}, \quad x \in \mathbb{X} \subset \mathbb{R}, \quad (4.17)$$

$$\mathcal{G}_k = \{g_k(x; \psi_{g_k}) \mid \psi_{g_k} \in \Psi_{g_k}\}, \quad y \in \mathbb{Y} \subset \mathbb{R}, \quad (4.18)$$

$$\mathcal{C}_k = \{c_0(u, v; \phi_k) \mid \phi_k \in \Phi_k\}, \quad (u, v) \in [0, 1]^2, \quad (4.19)$$

$$\begin{aligned} \mathcal{H}_k &= \{h_k(x, y) = f_k(x)g_k(y)c_k(F_k(x), G_k(y)) \\ &\quad \mid f_k \in \mathcal{F}_k, g_k \in \mathcal{G}_k, c_k \in \mathcal{C}_k\}, \end{aligned} \quad (4.20)$$

where, \mathbb{X} and \mathbb{Y} are closed with $x_L = \inf \mathbb{X}$, $y_L = \inf \mathbb{Y}$, $F_k(x) = \int_{x_L}^x f_k(x')dx'$ and $G_k(y) = \int_{y_L}^y g_k(y')dy'$.

A joint model \mathcal{H}_0 is nested in \mathcal{H}_1 if and only if marginal and copula models are both nested, i.e., \mathcal{F}_0 is nested in \mathcal{F}_1 , \mathcal{G}_0 is nested in \mathcal{G}_1 and \mathcal{C}_0 is nested in \mathcal{C}_1 .

Proof. We need to prove both “if” and “only if” parts.

Nested marginal models \Rightarrow nested joint model: It is easy to see that if $\mathcal{F}_0 \subset \mathcal{F}_1$, $\mathcal{G}_0 \subset \mathcal{G}_1$, and $\mathcal{C}_0 \subset \mathcal{C}_1$ then any product $f_0 g_0 c_0 \in \mathcal{H}_0$ is also contained in \mathcal{H}_1 , and hence $\mathcal{H}_0 \subset \mathcal{H}_1 \forall (x, y) \in \mathbb{X} \times \mathbb{Y}$.

Converse: For $k = 0, 1$, we first define the CDF models

$$\mathcal{H}'_k = \{H_k(x, y) = \int_{x_L}^x \int_{y_L}^y h_k(x', y')dx'dy' \mid h_k \in \mathcal{H}_k\},$$

$$\mathcal{C}'_k = \{C_k(u, v) = \int_0^u \int_0^v c_k(u', v')du'dv' \mid c_k \in \mathcal{C}_k\}.$$

Then, since $\mathcal{H}_0 \subset \mathcal{H}_1 \forall (x, y) \in \mathbb{X} \times \mathbb{Y}$, we have $\mathcal{H}'_0 \subset \mathcal{H}'_1$. Using (3.6) from Theorem 3.1, $\{C_0(F_0(x), G_0(y); \phi_0) \mid \phi_0 \in \Phi_0\} \subset \{C_1(F_1(x), G_1(y); \phi_1) \mid \phi_1 \in \Phi_1\}$. Since (a) \mathbb{X} and \mathbb{Y} are closed, and (b) F_k, G_k are continuous, $\mathcal{C}'_0 \subset \mathcal{C}'_1 \forall (u, v) \in [0, 1]^2$. Consequently $\mathcal{C}_0 \subset \mathcal{C}_1$, i.e., $c_0 \in \mathcal{C}_0 \Rightarrow c_0 \in \mathcal{C}_1$.

That $\mathcal{F}_0 \subset \mathcal{F}_1$ and $\mathcal{G}_0 \subset \mathcal{G}_1$ remains to be proved. We prove this by contradiction. Three contradictory cases exist: (a) $\mathcal{F}_0 \not\subset \mathcal{F}_1$ and $\mathcal{G}_0 \subset \mathcal{G}_1$; (b) $\mathcal{F}_0 \subset \mathcal{F}_1$ and $\mathcal{G}_0 \not\subset \mathcal{G}_1$; and (c) $\mathcal{F}_0 \not\subset \mathcal{F}_1$ and $\mathcal{G}_0 \not\subset \mathcal{G}_1$. We provide detailed arguments to show that case (c) is not

possible; cases (a) and (b) can be disproved using arguments similar to those presented below.

Assume $\exists \bar{f} \in \mathcal{F}_0 \not\subset \mathcal{F}_1$ and $\bar{g} \in \mathcal{G}_0 \not\subset \mathcal{G}_1$ so that $\bar{h}(x, y) = \bar{f}(x)\bar{g}(y)c_0(\bar{F}(x), \bar{G}(y))$, where $\bar{h}(x, y) \in \mathcal{H}_0 \subset \mathcal{H}_1$ (i.e., $\bar{h}(x, y) \in \mathcal{H}_1$). Therefore, following (4.20) for $k = 1$, there must exist marginal densities $\bar{f}_1 \neq \bar{f}$, $\bar{g}_1 \neq \bar{g}$ such that $\bar{f}_1 \in \mathcal{F}_1$, $\bar{g}_1 \in \mathcal{G}_1$ and $\bar{h}(x, y) = \bar{f}_1(x)\bar{g}_1(y)c_0(\bar{F}_1(x), \bar{G}_1(y))$. Thus, a single copula c_0 is associated with the same joint density $\bar{h}(x, y)$ for distinct marginal pairs $\{\bar{f}(x), \bar{g}(y)\}$ and $\{\bar{f}_1(x), \bar{g}_1(y)\}$. That is,

$$\bar{f}(x) \neq \bar{f}_1(x) \text{ contradicts } \bar{f}(x) = \int_{\mathbb{Y}} h(x, y) dy = \bar{f}_1(x) \text{ and}$$

$$\bar{g}(y) \neq \bar{g}_1(y) \text{ contradicts } \bar{g}(y) = \int_{\mathbb{X}} h(x, y) dx = \bar{g}_1(y),$$

implying, that $\nexists \bar{f} \in \mathcal{F}_0 \not\subset \mathcal{F}_1$, $\bar{g} \in \mathcal{G}_0 \not\subset \mathcal{G}_1$. Hence, $\mathcal{H}_0 \subset \mathcal{H}_1 \Rightarrow \mathcal{F}_0 \subset \mathcal{F}_1, \mathcal{G}_0 \subset \mathcal{G}_1$ and $\mathcal{C}_0 \subset \mathcal{C}_1 \forall (x, y) \in \mathbb{X} \times \mathbb{Y}$ \square

Lemma 4.1 allows us to state and prove chi-square convergence in distribution for T_{GLR} through Proposition 4.1.

Proposition 4.1. *Suppose that the hypothesis testing problem in (4.8) is specified as follows:*

- (i) *The joint distribution, $f_{\mathbf{X}}^1$, and parameter space under H_1 are not known since c_1 is unknown. The p.d.f. $f_{\mathbf{X}}^*$ and corresponding parameter space, Θ_{c^*} , are determined as an outcome of the copula selection process in (4.14).*
- (ii) *\mathbf{X} is well specified in $\{f_{\mathbf{X}}^0\} \cup \{f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta_{c^*}\}$.*
- (iii) *There may exist marginal parameters which are fixed and the same for both hypotheses. Denote the parameter subspace containing only free parameters as Θ'_{c^*} . Then, $\dim(\Theta_{c^*}) - \dim(\Theta'_{c^*}) = \nu \geq 0$.*
- (iv) *The copula model under H_0 is nested in the copula model selected under H_1 . That is, for known ϕ_{c_0} such that $\mathcal{C}_0 = \{c_0(u, v; \phi_{c_0})\}$, and for $\mathcal{C}^* = \{c^*(u, v; \phi_{c^*}) \mid \phi_{c^*} \in \Phi_{c^*}\}$, $\mathcal{C}_0 \subset \mathcal{C}^* \forall (u, v) \in [0, 1]^2$.*

Additionally, for sensor i , and hypothesis k , $\psi_{ik} \in \Psi_{ik} \subset \Psi$,

$$\begin{aligned} \Psi = \{[\alpha, \beta, \gamma, \delta]^\top \mid \alpha \in [\varepsilon, 1) \cup (1, 2), \varepsilon > 0; \\ |\beta| < \min(\alpha, 2 - \alpha); \gamma \in (0, \infty); \delta \in \mathbb{R}\}. \end{aligned} \quad (4.21)$$

Then, under H_0 , as $N \rightarrow \infty$,

$$2T_{\text{GLR}} \xrightarrow{D} \chi_{\mu-\nu}^2, \quad (4.22)$$

where, $\mu = \dim(\Theta_{c^*}) > \nu$ and $\chi_{\mu-\nu}^2$ is a chi-square random variable with $\mu - \nu$ degrees of freedom.

Proof. To prove χ^2 convergence, as seen in (4.22), we essentially invoke Wilks theorem (WT) [106]. However, to apply WT, we first need to prove that (a) the joint null model $\mathcal{H}_0 = \{f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}_0)\}$ is nested in the joint model $\mathcal{H}_1 = \{f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta_{c^*}\}$, and (b) that the parameter estimates obtained using the IFM method are asymptotically normal. The marginal models under the null hypothesis are nested for each $i = 1, 2$, i.e., $\{f_{X_i}(x_{ij}; \boldsymbol{\psi}_i) \mid \boldsymbol{\psi}_i = \boldsymbol{\psi}_{i0}\} \subset \{f_{X_i}(x_{ij}; \boldsymbol{\psi}_i) \mid \boldsymbol{\psi}_i \in \Psi_i\}$. Since both marginals and copula models are nested, from Lemma 4.1, $\mathcal{H}_0 \subset \mathcal{H}_1$.

We use primes to denote both, free parameters, which need to be estimated, and their corresponding subspaces. Thus, $\boldsymbol{\psi}'_i$ denotes the vector of free marginal parameters, for the i -th sensor, contained in the subspace Ψ'_i . Given (4.21),

$$\sqrt{N}(\widehat{\boldsymbol{\psi}}'_i - \boldsymbol{\psi}'_i) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\psi}'_i)), \quad (4.23)$$

where² $\mathcal{I}(\boldsymbol{\psi}'_i)$ is the corresponding Fisher information matrix evaluated at the true value, $\boldsymbol{\psi}'_i$ [26]. For marginal estimates, which are asymptotically normal, as in (4.23),

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}'_* - \boldsymbol{\theta}'_*) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathcal{G}^{-1}(\boldsymbol{\theta}'_*)), \quad (4.24)$$

² $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a normal random vector with mean \mathbf{m} , covariance \mathbf{C} .

where $\mathcal{G}(\boldsymbol{\theta}'_*)$ is the Godambe information matrix [41] evaluated at the true value, $\boldsymbol{\theta}'_*$. Since $\mathcal{H}_0 \subset \mathcal{H}_1$, and $\hat{\boldsymbol{\theta}}'_*$ is asymptotically normal WT yields (4.22). \square

Remark. In the proof of Proposition 4.1, the restriction on Ψ_i , as given in (4.21), is used to assert the asymptotic normality of $\hat{\boldsymbol{\psi}}_i$, under H_1 . We additionally require that (4.21) be true under H_0 to ensure that the two hypotheses are nested.

Proposition 4.1 implies that we can express the probability of false alarm as a function of the threshold, η , using the χ^2 CDF, F_{χ^2} , i.e.,

$$P_F(\eta) = 1 - F_{\chi^2}(2\eta; \mu - \nu).$$

Thus, for $P_F = a$, the detector threshold can be designed as,

$$\eta_a = 0.5 F_{\chi^2}^{-1}(1 - a; \mu - \nu). \quad (4.25)$$

Using (4.24), under the conditions listed in Proposition 4.1, the asymptotic convergence of the test-statistic under H_1 is an extension of Wald's result [101],

$$2 T_{\text{GLR}} \xrightarrow{D} \bar{\chi}_{\mu-\nu}^2 \left((\boldsymbol{\theta}'_{c^*} - \boldsymbol{\theta}'_0)^\top \mathcal{G}(\boldsymbol{\theta}'_{c^*}) (\boldsymbol{\theta}'_{c^*} - \boldsymbol{\theta}'_0) \right), \quad (4.26)$$

where $\bar{\chi}_r^2(p)$ denotes a non-central chi-square distribution with r degrees of freedom, and non-centrality parameter, p . The vector $\boldsymbol{\theta}'_0$ contains the parameters under H_0 ; it is a point in the $\mu - \nu$ dimensional subspace $\Theta'_{c_0} = \Psi'_1 \times \Psi'_2 \times \Phi_{c_0}$.

4.3.2 Non-nested models

We now consider the case when \mathcal{H}_0 is not nested in \mathcal{H}_1 . The general class of non-nested hypothesis testing problems was first considered by Cox [20, 21]. Subsequently, White [105] analyzed the problem to establish the regularity conditions under which Cox's proposed test

is valid. Vuong [100] has generalized this work, by defining precise forms of model nesting, and deriving a hypothesis testing based model selection scheme. These formulations consider a composite null hypothesis. We, however, formulate a simple null hypothesis. We derive the distribution of the test-statistic under H_0 and observe that it has a form which is similar to previously derived results [21, 74, 100, 105].

Non-nestedness implies that the p.d.f. under H_1 can have a different functional form as compared to the p.d.f. under H_0 . Consequently, though observations may be generated by \mathcal{H}_0 , likelihood maximization, in T_{GLR} , is carried out for a function in \mathcal{H}_1 . Hence, asymptotic convergence of the MLE to a pseudo-true value needs to be defined.

Definition 4.3 (Pseudo-true value and QMLE [74]). Suppose a DGP $\mathcal{H}_{\text{DGP}} = \{h(\mathbf{z}; \boldsymbol{\theta}_h) \mid \boldsymbol{\theta}_h = \boldsymbol{\theta}_{h0} \in \Theta_h\}$ is defined over the random vector $\mathbf{Z} \in \mathbb{Z} \subset \mathbb{R}^n$. The pseudo-true value for a model $\mathcal{G} = \{g(\mathbf{z}; \boldsymbol{\theta}_g) \mid \boldsymbol{\theta}_g \in \Theta_g\}$ is then defined as that value of $\boldsymbol{\theta}_g$ which minimizes the Kullback-Leibler divergence (KLD), D_{KL} , between h and g . That is,

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_g &\triangleq \arg \min_{\boldsymbol{\theta}_g \in \Theta_g} D_{\text{KL}}(h(\mathbf{z}; \boldsymbol{\theta}_{h0}) \parallel g(\mathbf{z}; \boldsymbol{\theta}_g)) \\ &= \arg \sup_{\boldsymbol{\theta}_g \in \Theta_g} \mathbb{E}_h \{ \log g(\mathbf{z}; \boldsymbol{\theta}_g) \}, \end{aligned} \quad (4.27)$$

where \mathbb{E}_h is the expectation under h . For the N -sample IID sequence $\{\mathbf{z}_j\}_{j=1}^N$, the quasi-maximum likelihood estimate (QMLE) is the sample estimate,

$$\hat{\boldsymbol{\theta}}_{g,N} = \arg \sup_{\boldsymbol{\theta}_g \in \Theta_g} (1/N) \sum_{j=1}^N \log g(\mathbf{z}_j; \boldsymbol{\theta}_g). \quad (4.28)$$

Under mild regularity conditions, it can be shown that $\hat{\boldsymbol{\theta}}_{g,N}$ exists and is a strongly consistent estimator of $\tilde{\boldsymbol{\theta}}_g$, i.e., $\hat{\boldsymbol{\theta}}_{g,N} \xrightarrow{a.s.} \tilde{\boldsymbol{\theta}}_g$ [104], and, therefore, $\hat{\boldsymbol{\theta}}_{g,N} \xrightarrow{P} \tilde{\boldsymbol{\theta}}_g$. Using the consistency of the QMLE, we next prove that, asymptotically, the model selection scheme of (4.14) will select the true copula.

Proposition 4.2. *If $\{c(u_1, u_2; \phi_c) \mid \phi_c \in \Phi_c\} \in \mathcal{C}$ is the copula model corresponding to the DGP, the selection process in (4.14) will select the true copula, c , asymptotically in N .*

Proof. Consider an arbitrary indexing set, \mathcal{M} , for the copula library, \mathcal{C} . Suppose

$$\{c_m(u_1, u_2; \phi_{c_m}) \mid \phi_{c_m} \in \Phi_{c_m}\} \in \mathcal{C}$$

such that $c_m \neq c$ for some $m \in \mathcal{M}$. Recall that (4.14) maximizes $c_m(\cdot; \phi_{c_m})$ over ϕ_{c_m} for every $c_m \in \mathcal{C}$, and selects that copula which has the maximum likelihood over all $m \in \mathcal{M}$. Also, likelihood maximization does not change under normalization by N , i.e., for $\ell_{c,N}(\phi_c) = \sum_{j=1}^N \log c(u_{1j}, u_{2j}; \phi_c)$,

$$\hat{\phi}_{c,N} = \arg \sup_{\phi_c \in \Phi_c} \ell_{c,N}(\phi_c) = \arg \sup_{\phi_c \in \Phi_c} N^{-1} \ell_{c,N}(\phi_c).$$

A similar observation holds for c_m . Although the true parameter value of c is unknown *a priori*, it exists and is denoted as ϕ'_c . Since $\hat{\phi}_{c,N} \xrightarrow{P} \phi'_c$ and $\hat{\phi}_{c_m,N} \xrightarrow{P} \tilde{\phi}_{c_m}$, we can write³:

$$\text{plim}_{N \rightarrow \infty} \sum_{j=1}^N [\log c(\cdot; \hat{\phi}_{c,N}) - \log c_m(\cdot; \hat{\phi}_{c_m,N})] / N \quad (4.29)$$

$$= \mathbb{E}_c[\log \{c(u_1, u_2; \phi'_c) / c_m(u_1, u_2; \tilde{\phi}_{c_m})\}] \quad (4.30)$$

$$= D_{\text{KL}}(c \parallel c_m) > 0 \quad (4.31)$$

where \mathbb{E}_c is the expectation under c and (4.30) is a consequence of the law of large numbers. In (4.31), $D_{\text{KL}}(c \parallel c_m)$ denotes the KLD between c and c_m ; the inequality is strict since $c \neq c_m$. From (4.29) and (4.31),

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \log c(\cdot; \hat{\phi}_{c,N}) > \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \log c_m(\cdot; \hat{\phi}_{c_m,N}).$$

³If $X_N \xrightarrow{P} X$ as $N \rightarrow \infty$ we can equivalently write $X = \text{plim}_{N \rightarrow \infty} X_N$.

The above arguments hold true for all $c_m \in \mathcal{C}$ distinct from c . Therefore, using (4.14), c will be selected as $N \rightarrow \infty$ \square

Proposition 4.2 is significant because it implies that, when the data are generated under H_0 , if c_0 is contained in \mathcal{C} , copula selection will always (asymptotically) select c_0 as c^* . In effect, there are very specific cases when T_{GLR} , under H_0 , is evaluated from p.d.f.s contained in non-nested models:

1. marginals are nested, but we know the function c_1 *a priori*;
2. marginals come from the same family (α -stable in this paper), but not nested because the marginal parameters are defined over disjoint subspaces over the hypotheses.

For the first case (nested α -stable marginals and non-nested copulas), since the functional form of the copula under H_1 is known, \mathcal{C} contains only c_1 , and, trivially, $c^* = c_1$. Hence, there is no model selection component to the detection problem. Irrespective of the true hypothesis, the marginal parameter estimates, $\hat{\psi}_i$, will converge to the true value, ψ_i and, therefore, $u_{ij}(\hat{\psi}_i)$ will be asymptotically uniform. Thus, $\hat{\phi}_{c_1}$, obtained by maximizing $c_1(u_{1j}(\hat{\psi}_1), u_{2j}(\hat{\psi}_2); \phi_{c_1})$, will converge to $\tilde{\phi}_{c_1}$ and, therefore, we can use the IFM method for copula parameter estimation.

The second case, with non-nested marginals, will occur when the problem is defined such that H_0 is specified for a marginal parameter that cannot be achieved by likelihood maximization under H_1 . This includes the case where H_0 is defined as one of the points not allowed for α -stable ML estimation (see (4.21)), e.g., when dependent Cauchy distributed marginals under H_0 ($\alpha_i = 1$) are tested against the composite alternative of dependent, stable distributed marginals. Alternatively, there may exist additional knowledge about H_1 that indicates a restricted marginal parameter range; $\hat{\psi}_i$ is then a range-restricted MLE. When the marginals are not nested, $u_{ij}(\hat{\psi}_i)$ is not asymptotically distributed as $\mathcal{U}(0, 1)$, and the IFM estimate, $\hat{\phi}_{c^*}$, is

not consistent. In this case, we use the nonparametric empirical CDF (ECDF),

$$\widehat{F}_{X_i}(t) = (1/N) \sum_j \mathbb{1}\{x_{ij} < t\}. \quad (4.32)$$

In (4.12), the estimate $\hat{u}_{ij} = \widehat{F}_{X_i}(x_{ij})$ is used in place of $u_{ij}(\widehat{\psi}_i)$. Since $\widehat{F}_{X_i} \xrightarrow{P} F_{X_i}$, it can be shown that a two-step procedure, using \hat{u}_{ij} instead of $u_{ij}(\widehat{\psi}_i)$, also leads to a consistent estimate of the copula parameter [41].

When the data are generated by H_1 , both, $u_{ij}(\widehat{\psi}_i)$ and \hat{u}_{ij} , are asymptotically uniform, and as a result $\widehat{\phi}_{c^*}$ is consistent. Thus, Proposition 4.1 and (4.26) hold if \hat{u}_{ij} is used instead of $u_{ij}(\widehat{\psi}_i)$. In Proposition 4.3, we establish the asymptotic distribution of T_{GLR} under H_0 for non-nested hypotheses.

Proposition 4.3. *For the formulation in (4.8), the null and alternative p.d.f. models are \mathcal{H}_0 , \mathcal{H}_1 , respectively, so that $\mathcal{H}_0 \not\subset \mathcal{H}_1$. Denote the pseudo-true value of θ under H_1 as $\widetilde{\theta}_1$ so that $\tilde{f}_X^1 = f_X(x_j; \widetilde{\theta}_1)$. Under H_0 ,*

$$\sqrt{N} \left(T_{\text{GLR}}/N + D_{\text{KL}}(f_X^0 \| \tilde{f}_X^1) \right) \xrightarrow{D} \mathcal{N}(0, \tilde{w}^2), \text{ where} \quad (4.33)$$

$$\tilde{w}^2 = \mathbb{E}_0[\{\log f_X^0 / \tilde{f}_X^1\}^2] - \{\mathbb{E}_0[\log f_X^0 / \tilde{f}_X^1]\}^2 \quad (4.34)$$

and \mathbb{E}_0 is the expectation under f_X^0 .

Proof. Let the log-likelihood under H_0 be $\ell_N(\theta_0) = \sum_j \log f_X(x_j; \theta_0)$, $\theta_0 \in \Theta_{c_0} \subset \mathbb{R}^{p_0}$. Under (composite) H_1 as $\ell_N(\widehat{\theta}_*) = \sum_j \log f_X(x_j; \widehat{\theta}_*)$, $\widehat{\theta}_* \in \Theta_{c^*} \subset \mathbb{R}^{p_1}$. Expanding $\ell_N(\widetilde{\theta}_1) = \sum_j \log f_X(x_j; \widetilde{\theta}_1)$, $\widetilde{\theta}_1 \in \mathbb{R}^{p_1}$, about $\widehat{\theta}_*$, using the mean-value form for the remainder in Taylor's theorem, we obtain

$$\ell_N(\widetilde{\theta}_1) = \ell_N(\widehat{\theta}_*) + [\nabla_{\theta} \ell_N(\widehat{\theta}_*)]^T (\widetilde{\theta}_1 - \widehat{\theta}_*) + \frac{1}{2} (\widetilde{\theta}_1 - \widehat{\theta}_*)^T [\nabla_{\theta}^2 \ell_N(\bar{\theta})] (\widetilde{\theta}_1 - \widehat{\theta}_*), \quad (4.35)$$

where $\bar{\theta}$ lies on the segment joining $\widetilde{\theta}_1$ and $\widehat{\theta}_*$. As a consequence of likelihood maximiza-

tion, $\nabla_{\boldsymbol{\theta}} \ell_N(\widehat{\boldsymbol{\theta}}_*) = \mathbf{0}$. The law of large numbers implies that $(1/N) \nabla_{\boldsymbol{\theta}}^2 \ell_N(\widehat{\boldsymbol{\theta}}_*) \xrightarrow{P} \mathbb{E}_0[\nabla_{\boldsymbol{\theta}}^2 \tilde{f}_{\mathbf{X}}^1]$, where the expectation is under H_0 since it corresponds to the hypothesis under which the data are being generated. This convergence also holds for the likelihood evaluated at $\bar{\boldsymbol{\theta}}$ as $(\tilde{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_*) \xrightarrow{P} 0$ and, thus, the error term $(\tilde{\boldsymbol{\theta}}_1 - \bar{\boldsymbol{\theta}})$ converges similarly. Therefore, we can write $(1/N) \nabla_{\boldsymbol{\theta}}^2 \ell_N(\bar{\boldsymbol{\theta}}) = \mathbb{E}_0[\nabla_{\boldsymbol{\theta}}^2 \tilde{f}_{\mathbf{X}}^1] + \mathbf{Y}_N$, where \mathbf{Y}_N is a $p_1 \times p_1$ matrix such that each element is a $o_p(1)$ random variable⁴. Let $\mathbf{A}_1 = \mathbb{E}_0[\nabla_{\boldsymbol{\theta}}^2 \tilde{f}_{\mathbf{X}}^1]$ and $\boldsymbol{\theta}_{1,N}^\epsilon = \tilde{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_*$ so that (4.35) becomes

$$\ell_N(\tilde{\boldsymbol{\theta}}_1) = \ell_N(\widehat{\boldsymbol{\theta}}_*) + \underbrace{(N/2) \boldsymbol{\theta}_{1,N}^{\epsilon \top} (\mathbf{A}_1 + \mathbf{Y}_N) \boldsymbol{\theta}_{1,N}^\epsilon}_{=J_N}. \quad (4.36)$$

where,

$$J_N = \frac{N}{2} \boldsymbol{\theta}_{1,N}^{\epsilon \top} \mathbf{A}_1 \boldsymbol{\theta}_{1,N}^\epsilon + \frac{1}{2} (\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon)^\top \mathbf{Y}_N (\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon).$$

Also, $\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon \xrightarrow{D} \mathcal{N}(0, \mathbf{A}_1^{-1} \mathbf{B}_1 \mathbf{A}_1^{-1})$, $\mathbf{B}_1 = \mathbb{E}_0[\nabla_{\boldsymbol{\theta}} \tilde{f}_{\mathbf{X}}^1 \cdot \nabla_{\boldsymbol{\theta}}^\top \tilde{f}_{\mathbf{X}}^1]$ (see [104, Theorem 3.2]). Each element of $(\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon)$ converges in distribution to a normal random variable, and hence, is bounded in probability⁵ (see Theorem 2.3.2 in [57]). Lemma 2.3.1 and Theorem 2.1.3 from [57] prove that⁶ $(\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon)^\top \mathbf{Y}_N (\sqrt{N} \boldsymbol{\theta}_{1,N}^\epsilon)$ is $o_P(1)$. Set $\tilde{T}_N = \ell_N(\tilde{\boldsymbol{\theta}}_1) - \ell_N(\boldsymbol{\theta}_0)$. Note that $T_{\text{GLR},N} \equiv T_{\text{GLR}} = \ell_N(\widehat{\boldsymbol{\theta}}_*) - \ell_N(\boldsymbol{\theta}_0)$. Thus,

$$T_{\text{GLR},N} = \tilde{T}_N - (N/2) \boldsymbol{\theta}_{1,N}^{\epsilon \top} \mathbf{A}_1 \boldsymbol{\theta}_{1,N}^\epsilon + o_P(1) \quad (4.37)$$

$$\begin{aligned} &\Rightarrow \sqrt{N} (T_{\text{GLR},N}/N + \mathbb{E}_0[\log f_{\mathbf{X}}^0 / \tilde{f}_{\mathbf{X}}^1]) \\ &= -\sqrt{N} (-\tilde{T}_N/N - \mathbb{E}_0[\log f_{\mathbf{X}}^0 / \tilde{f}_{\mathbf{X}}^1]) \\ &\quad - (\sqrt{N} \boldsymbol{\theta}_{1,N}^{\epsilon \top} \mathbf{A}_1 \boldsymbol{\theta}_{1,N}^\epsilon)/2 + (1/\sqrt{N}) o_P(1). \end{aligned} \quad (4.38)$$

\mathbf{A}_1 does not depend on N and $\boldsymbol{\theta}_{1,N}^\epsilon$ is $o_P(1)$. Consequently, using Lemma 2.3.1 from [57],

⁴A random variable X_N is $o_P(T_N)$ if $(X_N/T_N) \xrightarrow{P} 0$ as $N \rightarrow \infty$.

⁵ X_N is said to be bounded in probability (denoted $O_P(1)$) iff for every $\epsilon > 0 \exists B_\epsilon < \infty$ and N_ϵ such that $\Pr[|X_N| \leq B_\epsilon] > 1 - \epsilon \forall N \geq N_\epsilon$.

⁶The lemma and two theorems are informally stated as follows (see [57] for more details). *Theorem 2.3.2:* $X_N \xrightarrow{D} X \Rightarrow X_N = O_P(1)$; *Lemma 2.3.1:* $o_P(1) \cdot O_P(1) = o_P(1)$; *Theorem 2.1.3:* $o_P(1) + o_P(1) = o_P(1)$.

$\sqrt{N}\boldsymbol{\theta}_{1,N}^\epsilon{}^\top \mathbf{A}_1 \boldsymbol{\theta}_{1,N}^\epsilon$ is $o_P(1)$. Therefore,

$$\sqrt{N}\left(\frac{T_{\text{GLR},N}}{N} + D_{\text{KL}}(f_{\mathbf{X}}^0 \| \tilde{f}_{\mathbf{X}}^1)\right) = o_P(1) - \sqrt{N}\{(1/N)\sum_j \log f_{\mathbf{X}}^0 / \tilde{f}_{\mathbf{X}}^1 - \mathbb{E}_0[\log f_{\mathbf{X}}^0 / \tilde{f}_{\mathbf{X}}^1]\}.$$

The second term in the RHS contains the sample mean of $\log f_{\mathbf{X}}^0 / \tilde{f}_{\mathbf{X}}^1$ and its expectation; apply CLT to get (4.34). \square

Although \tilde{w}^2 depends on the pseudo-true value under H_1 , it can be consistently estimated from the observations as

$$\hat{w}^2 = (1/N)\{\sum_{j=1}^N T_{\text{GLR},j}^2 - T_{\text{GLR}}^2\}, \quad (4.39)$$

where $T_{\text{GLR},j} = \log f_{\mathbf{X}}^* / f_{\mathbf{X}}^0$ [100]. The asymptotic distribution under H_1 can be proved in a manner similar to Proposition 4.3.

Proposition 4.4. *For the hypothesis testing problem in Proposition 4.3, under H_1 , as $N \rightarrow \infty$,*

$$\{T_{\text{GLR}}/N - D_{\text{KL}}(f_{\mathbf{X}}^1 \| f_{\mathbf{X}}^0)\}/(\hat{w}/\sqrt{N}) \xrightarrow{D} \mathcal{N}(0, 1) \quad (4.40)$$

Proof. We proceed along similar lines as Proposition 4.3, but by replacing \mathbb{E}_0 by \mathbb{E}_1 , i.e., the expectation under H_1 . Note that since the data are generated under H_1 , we use $\boldsymbol{\theta}_1$ instead of $\tilde{\boldsymbol{\theta}}_1$. Denote the variance of T_{GLR} under H_1 as

$$w_1^2 = \mathbb{E}_1[(\log f_{\mathbf{X}}^1 / f_{\mathbf{X}}^0)^2] - (\mathbb{E}_1[\log f_{\mathbf{X}}^1 / f_{\mathbf{X}}^0])^2.$$

The limit in (4.40) follows as $\hat{w}^2 \xrightarrow{P} w_1^2$. \square

Equation (4.34) implies that determining the detector threshold requires knowledge of the Kullback-Leibler divergence evaluated at the (pseudo-) true values of the distribution parame-

ters under each hypothesis. Specifically, for $P_F = a$,

$$\eta_a = \sqrt{N}\hat{w}F_{\mathcal{N}}^{-1}(1 - a) - ND_{\text{KL}}(f_{\mathbf{X}}^0 \| f_{\mathbf{X}}^1),$$

where $F_{\mathcal{N}}^{-1}$ is the inverse CDF of the standard normal distribution and \hat{w} is obtained from data generated under H_0 . Numerical methods may be employed to compute D_{KL} for a two-sensor formulation; however, it is easy to see that this is not scalable for a multi-sensor formulation. In such scenarios, existing data-driven approaches such as bootstrapping, or extreme-value theory (EVT) based distribution fitting of T_{GLR} under H_0 yield reliable approximations of η_a . The latter approach based on EVT is especially attractive for detection with α -stable distributions. This is because EVT indicates that the null distribution of likelihood ratio test statistics is distributed asymptotically in the tails as a generalized Pareto distribution. This behavior would be even more evident for stable-distributed observations, since such random variables exhibit Pareto tails [27]. Ozturk et al. [72] provide a detailed treatment of estimating detector threshold based on EVT.

4.4 Performance evaluation

In this section, we first illustrate the performance of the copula-based GLRT with the aid of simulated data from specifically constructed examples (Section 4.4.1). We address computational challenges and discuss footstep detection using the seismic sensor data described in Chapter 2, in Section 4.4.2 and Section 4.4.3, respectively.

4.4.1 Simulated examples

For each hypothesis, 50×10^5 pseudorandom sample pairs, representing dependent α -stable sensor measurements \mathbf{x}_j , are generated using MATLAB [61]. Estimates $\hat{\psi}_i$, $\hat{\phi}_{c^*}$ and \hat{u}_{ij} , and the test statistic, T_{GLR} , are computed for every distinct group of $N = 50$ samples. P_F and P_D

values are evaluated over the resultant 10^5 length sequence of T_{GLR} values.

We use Kendall's tau, τ , to quantify the dependence for a given copula, c , instead of directly specifying ϕ_c . This allows for a common basis for comparison across all examples being considered. A rank-based measure of dependence, τ is defined as the difference of probabilities of concordance and discordance. For the copula CDF, C , $\tau = 4 \mathbb{E}[C(u, v; \phi_c)] - 1$, and thus there exists a one-to-one relationship between τ and ϕ_c [65]. Analogous to ρ , $\tau \in [-1, 1]$: $\tau < 0$ and $\tau > 0$ indicate negative and positive dependence, respectively; independence implies that $\tau = 0$.

For the three examples considered below, in (4.41), (4.43) and (4.44), the parameter values, as listed, are used for data generation. While evaluating the performance of our proposed approach, boxed terms are assumed unknown for data processing, and are determined by estimation or model selection as a part of our detection methodology.

Example 4.1 (Dependent $S_\alpha S$ distributions).

$$\begin{aligned}
 H_0 : \quad X_1 &\sim \mathcal{S}(1.3, 0, 0.1, 0) & X_2 &\sim \mathcal{S}(1.5, 0, 0.1, 0) \\
 c_0 &= c_{\mathcal{N}}(\tau = 0.1) \\
 H_1 : \quad X_1 &\sim \mathcal{S}(1.3, 0, 0.1, \boxed{0.1}) & X_2 &\sim \mathcal{S}(1.5, 0, \boxed{0.2}, 0) \\
 c_1 &= \boxed{c_t(\tau = 0.5, \nu = 3)}
 \end{aligned} \tag{4.41}$$

Sensor 1 measures shift in mean, and Sensor 2 measures change in dispersion. The dependence under the null hypothesis is symmetric and is modeled using the Gaussian copula, $c_{\mathcal{N}}$. The dependence under H_1 is modeled using a t -copula, c_t , with $\nu = 3$ degrees of freedom (DoF). Both Gaussian and t copulas use the correlation coefficient, ρ , as the dependence parameter (see Table 4.2). Under H_0 , $\tau = 0.1 \Leftrightarrow \rho = 0.15$ indicates weaker dependence compared to $\tau = 0.5 \Leftrightarrow \rho = 0.7$ under H_1 . The empirical receiver operating characteristic (ROC) using T_{GLR} as the test-statistic is shown as the solid curve in Fig. 4.1. Recall that T_{GLR} , in (4.12), uses the copula selection procedure in (4.14). Often, one is tempted to assume that the dependence

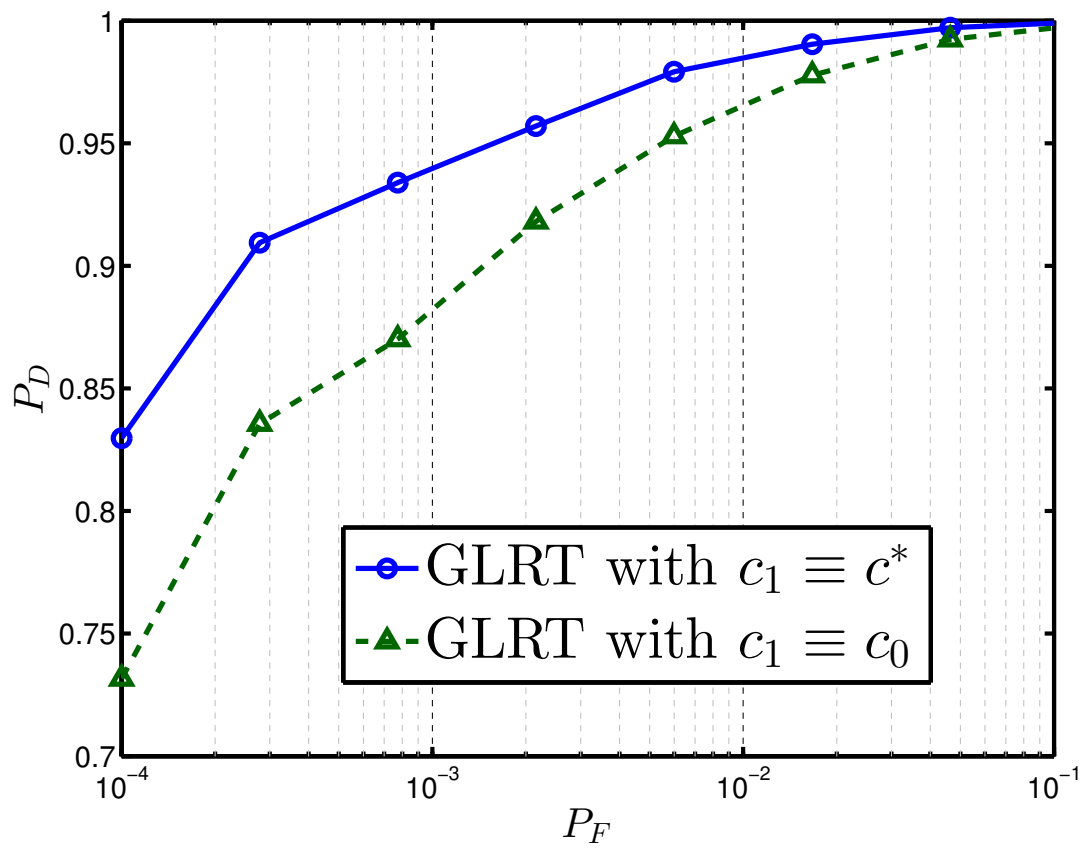


Fig. 4.1: ROC for Example 1: dependent $S\alpha S$ distributions.

model under the null hypothesis also prevails under the alternative hypothesis. This assumption implies that instead of using c^* , as would be obtained using (4.14), we use the Gaussian copula under H_1 so that the test-statistic is

$$T' = \sum_{j=1}^N \log\{f_{\mathbf{X}}(\mathbf{x}_j; [\hat{\delta}_1, \hat{\gamma}_2, \hat{\phi}_{c_N}]^T) / f_{\mathbf{X}}(\mathbf{x}_j; \boldsymbol{\theta}_0)\} \quad (4.42)$$

with $N = 50$. While the Gaussian copula under H_0 does not capture any tail dependence, the t -copula exhibits both lower and upper tail dependence. Lower DoF values indicate heavier dependence in the tails, i.e., extreme events co-occur with greater probability. As $\nu \rightarrow \infty$, a t -copula converges to a Gaussian copula. In that sense, ν controls the amount of tail dependence. As a consequence of mismodeling the copula, the tail dependence is inadequately characterized and the detector using T' suffers a 10% decrease in P_D for $P_F \leq 10^{-3}$. The ROC for T' is the dashed curve in Fig. 4.1.

Example 4.2 (Nearly normal distributions). As in Example 1, Sensor 1 and Sensor 2 measure mean and dispersion, respectively. The standard normal distribution is equivalent to $\mathcal{S}(2, 0, 0.5, 0)$. A tail index of 1.9 comes close to a normal distribution, but the tail still decays at a polynomial rate. The problem is setup as,

$$\begin{aligned} H_0 : \quad X_1 &\sim \mathcal{S}(1.9, 0, 1, 0) & X_2 &\sim \mathcal{S}(1.9, 0, 1, 0) \\ c_0 &= c_{\mathcal{N}}(\tau = 0.064) \\ H_1 : \quad X_1 &\sim \mathcal{S}(1.9, 0, 1, \boxed{1}) & X_2 &\sim \mathcal{S}(1.9, 0, \boxed{1.5}, 0) \\ c_1 &= \boxed{c_t(\tau = 0.41, \nu = 15)} \end{aligned} \quad (4.43)$$

The values of τ under H_0 and H_1 correspond (approximately) to correlation coefficient values of 0.1 and 0.6, respectively. At $\nu = 15$, the t -copula exhibits moderate to low tail dependence. Example 1 demonstrated the effect of assuming the null dependence model in the alternative hypothesis. A more egregious assumption is one of joint normality, as it represents

a case where, apart from the copula, the marginal model is also mismatched with respect to the DGP. If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\sigma_i^2 = 2\gamma_i^2$. The GLR statistic, T' , assumes joint normality under both hypotheses. For $H_0 : \mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{C}_0)$, $\boldsymbol{\mu}_0 = [0, 0]^\top$ and $\mathbf{C}_0 = \begin{bmatrix} 2 & 0.2 \\ 0.2 & 2 \end{bmatrix}$. Therefore,

$$T' = N \log |\hat{\mathbf{C}}| + \sum_{j=1}^N (\mathbf{x}_j - \hat{\boldsymbol{\mu}})^\top \hat{\mathbf{C}}^{-1} (\mathbf{x}_j - \hat{\boldsymbol{\mu}}) - \mathbf{x}_j^\top \mathbf{C}_0 \mathbf{x}_j$$

where $N = 50$, $\hat{\mathbf{C}} = \begin{bmatrix} 2 & \sqrt{2}\hat{\sigma}_2\hat{\rho} \\ \sqrt{2}\hat{\sigma}_2\hat{\rho} & \hat{\sigma}_2^2 \end{bmatrix}$ and $\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_1 \\ 0 \end{bmatrix}$.

Fig. 4.2 compares the ROCs using T_{GLR} and T' . The severe loss in P_D is clearly evident for the latter case.

Example 4.3 (Skewed and asymmetrically dependent distributions). The problem setup is as follows:

$$\begin{aligned} H_0 : \quad & X_1 \sim \mathcal{S}(1.5, 0, 0.1, 0) \quad X_2 \sim \mathcal{L}(0, 0.25) \\ & c_0 = 1 \text{ (Independent)} \\ H_1 : \quad & X_1 \sim \mathcal{S}(1.5, 0, 0.1, \boxed{0.1}) \quad X_2 \sim \mathcal{L}(0, \boxed{1}) \\ & c_1 = \boxed{c_{\text{Gu}}(\tau = 0.6)} \end{aligned} \tag{4.44}$$

where $\mathcal{L}(\delta, \gamma)$ is a Lévy distribution so that, $\mathcal{L}(\delta, \gamma) \equiv \mathcal{S}(0.5, 1, \gamma, \delta)$, and c_{Gu} represents a Gumbel copula. The Lévy distribution admits a closed-form p.d.f., so that for the j -th observation under hypothesis k ,

$$f_{X_2}(x_{2j}; \delta, \gamma_k) = \sqrt{\frac{\gamma_k}{2\pi}} \frac{\exp \left\{ -\gamma_k (2(x_{2j} - \delta))^{-1} \right\}}{(x_{2j} - \delta)^{3/2}}, \tag{4.45}$$

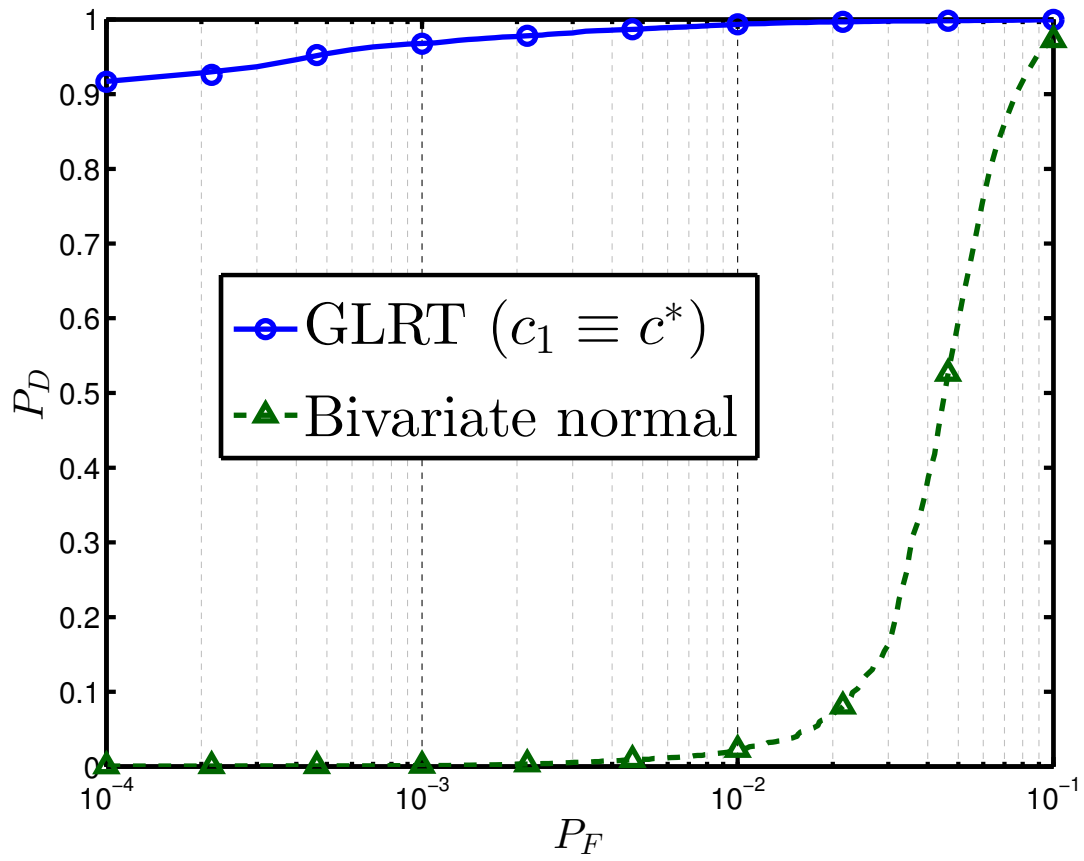


Fig. 4.2: ROC for Example 2: nearly normal distributions.

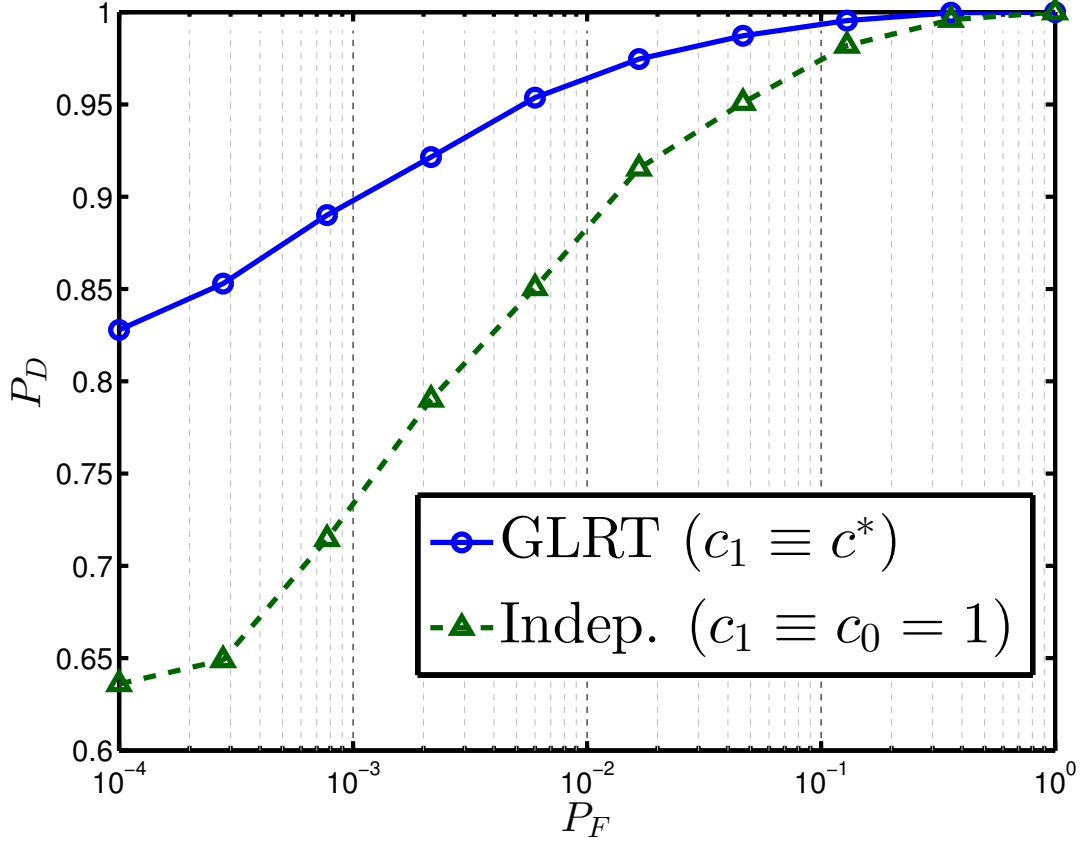


Fig. 4.3: ROC for Example 3: skewed and asymmetrically dependent distributions.

for $x_{2j} \in [\delta, \infty)$. For $\delta = 0$, the MLE for γ is,

$$\hat{\gamma} = N \left(\sum_{j=1}^N x_{2j}^{-1} \right)^{-1}.$$

The ROCs in Fig. 4.3 show that, for a given P_F value, the detector using T_{GLR} outperforms the detector assuming independence under H_1 (i.e., set the third term in (4.12) to 0).

Fig. 4.4 shows the contour plot of the joint density under H_1 from Example 3 and illustrates two key advantages of copula based modeling. While the observations from Sensor 1 are supported on \mathbb{R} , the observations from Sensor 2 are supported on $[0, \infty)$. Using the copula-based approach we are able to synthesize a valid joint p.d.f. from disparate marginals, which can capture the dependence in the tails. Secondly, skewed and nonlinear dependence is also

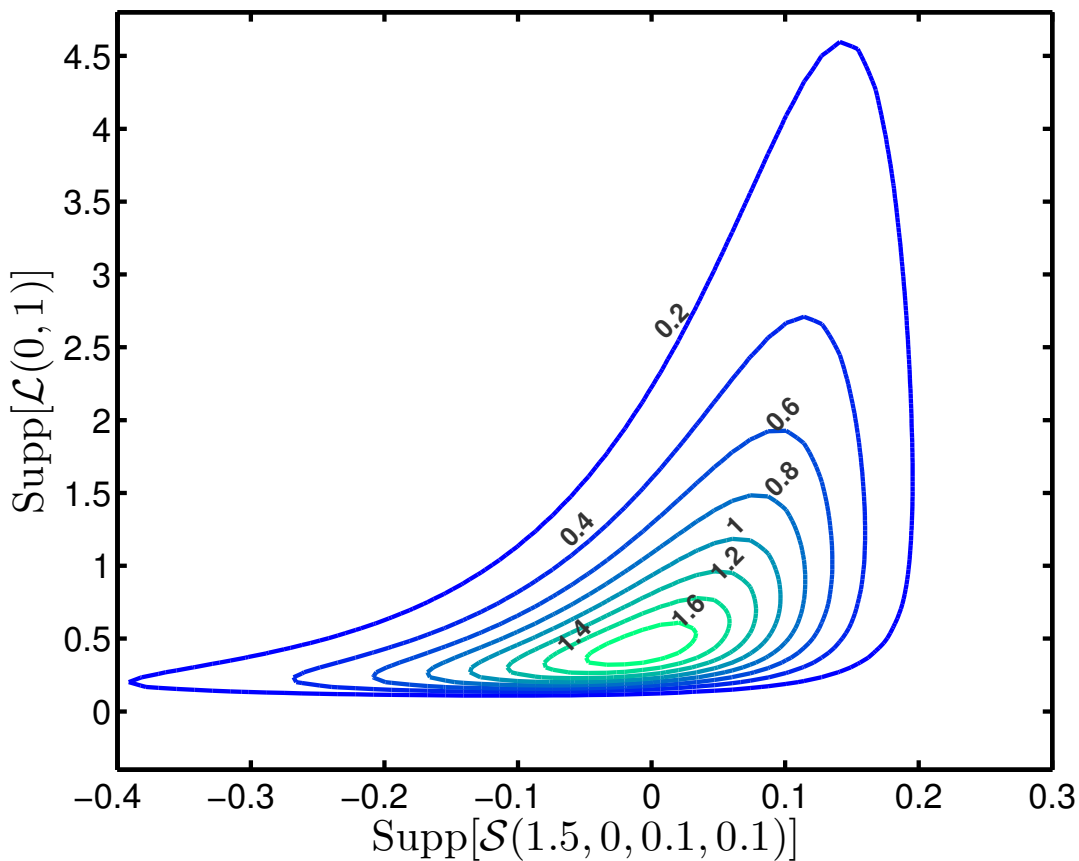


Fig. 4.4: Contour plot for f_X^1 from Example 3. The X and Y axes are the supports for the marginal densities of X_1 and X_2 , respectively.

adequately modeled; this may be contrasted with the typical concentric ellipses observed when using symmetric linear dependence models.

4.4.2 Computational considerations

In the examples discussed above, the hypotheses were constructed to reflect commonly observed scenarios in typical signal processing applications. For these cases, only the displacement or scale parameters need to be estimated. In many applications, however, MLE for the complete set of parameters is required and presents a significant computational burden.

Other than the number of parameters to be estimated, MLE is also constrained by (4.21). As an example, consider the case where a random variable $X \sim \mathcal{S}(1.3, 1, 0.2, 0)$. Let $\ell_X(\gamma)$

represent the log-likelihood of X , expressed as a function of γ over N samples. For $\alpha = 1.3, \beta = 1, |\beta| > \min(\alpha, 2 - \alpha)$ which violates (4.21). This affects $\ell_X(\gamma)$ even if we wish to estimate only γ : $\ell_X(\gamma)$ is unbounded and $\ell_X(\gamma) \rightarrow \infty$ as $\gamma \rightarrow 0$. The MLE $\hat{\gamma}_N$ will not converge to $\gamma = 0.2$ as $N \rightarrow \infty$.

As an alternative, iterative estimators such as the Koutrouvelis regression estimator (KRE) can be used [50]. The KRE is consistent and is computationally more efficient than the MLE. This is because they are derived from the characteristic function of a stable distribution. Furthermore, for certain cases, the KR estimates are also asymptotically normal [51]. This implies that the asymptotic distributions for T_{GLR} , derived in Section 4.3, typically hold for marginal estimates obtained using KRE as well. Perhaps counter-intuitively, when using skewed marginal distributions, we have observed in our experiments that improved detection performance may be obtained when the entire parameter set is estimated using KRE, when compared to using the MLE. This behavior may be observed even if the log-likelihood is maximized over a smaller subspace of unknown parameters.

For copula parameters, estimating Kendall's tau (τ) can be considered as a computationally efficient alternative to the MLE. Kendall's tau can be estimated non-parametrically. The estimates are consistent and since, for each copula in Table 4.2, there exists an invertible function ξ_c such that $\tau = \xi_c(\phi_c)$, we can obtain a copula parameter estimate, $\check{\phi}_c = \xi_c^{-1}(\hat{\tau})$, where

$$\hat{\tau} = (N_c - N_d)/(N_c + N_d) = (N_c - N_d)/\binom{N}{2}$$

for N_c concordant pairs and N_d discordant observation pairs⁷.

In the following subsection, the KRE and Kendall's tau based estimators are used to obtain parameter estimates for the log-likelihood ratios for footstep data collected in indoor environments. A footstep detection problem against a null “background” hypothesis is formulated.

⁷Observation pairs (x_{1j}, x_{2j}) and $(x_{1j'}, x_{2j'})$ are concordant if $(x_{1j} - x_{1j'})(y_{1j} - y_{1j'}) > 0$, and discordant if $(x_{1j} - x_{1j'})(y_{1j} - y_{1j'}) < 0$ [65].

Under the alternative “footstep” hypothesis, inter-sensor dependence is modeled using copulas. For the footstep detection problem, an additional element of uncertainty is that the *true* data generating copula is unknown and may not be contained in the copula library. The copula selection process, therefore, does not guarantee, even asymptotically, that the true copula will be chosen. However, the best model, in the KL divergence sense, will be selected. This is theoretically consistent with minimum description length (MDL) based approaches to model selection, especially for single parameter models. The CDF arguments for the copula under the alternative model is obtained using the empirical CDF in (4.32).

4.4.3 Footstep Detection

The preceding discussion considered simulated data. The remainder of this section discusses results obtained on applying the proposed detection scheme to the indoor seismic data described in Chapter 2. In this section, we, restrict our attention to the background and walking trials obtained from the two sensors at the “center” of the array.

Recall that each “walking” trial lasted approximately 12 s and the duration of the background data is 30 s. For the analysis in this section, the background data was split into a pilot (training) set and a test set. Each of the background and footstep datasets, were split into 1500 non-overlapping windows with 500 samples per window. The average walking pace is measured to be approximately 2 steps per second; therefore, for a 1 kHz sampling rate, a window length of 500 samples allows us to capture the dynamics of a footfall within one observation window. The data points in each window are used for parameter estimation and calculating a corresponding test-statistic. Consequently, 1500 decision windows are available over which P_F and P_D evaluations are made.

Data analysis for α -stable characterization

The background and footstep data, acquired from the geophone sensors, were analyzed in Chapter 2, in which we noted the significantly heavy-tailed nature of footstep data. In the following discussion we analyze the α -stable characterization of this data.

Specifically, we analyzed the α -stable fit for footstep data to contrast its behavior with respect to the background (null hypothesis) data. Table 4.3 lists the mean, median and standard error, median absolute deviation (MAD) statistics for the fitted α -stable parameters. These statistics are also shown for the Kendall's tau estimate of the data. The mean and median values are identical to the second decimal place for the α -stable parameters, especially for the background. A minor discrepancy exists between the mean and median values for τ , but τ for background data is rather small, and this difference does not affect our analysis and results. In the ensuing implementation of the detection algorithm, if there is a discrepancy between mean and median values, we use the median estimates as they are more robust. While τ values appear small for the footstep data, this must be taken in context with two points: (a) the Kendall's tau for the footstep data is an order of magnitude greater than the τ estimates for background, and (b) the τ estimates are affected by the interstitial periods of background that occur between two consecutive footfalls. When the footfall periods are isolated, τ estimates for these durations vary between 0.25 to 0.3.

Fig. 4.5 shows the scatter plot of 5000 randomly chosen observation pairs from the pilot background data along with a 99 % confidence ellipse for a bivariate $\mathcal{N}([0, 0]^T, \Sigma)$ distribution with covariance matrix

$$\Sigma \approx \begin{bmatrix} 0.58 & 0.009 \\ 0.009 & 0.05 \end{bmatrix}.$$

We notice that almost all points are encompassed within the confidence ellipse. The diagonal entries for Σ are obtained by using the relation $\sigma^2 = 2\gamma^2$ for a normal distribution, and applying it to the γ values in Table 4.3 corresponding to background data. The off-diagonal covariance

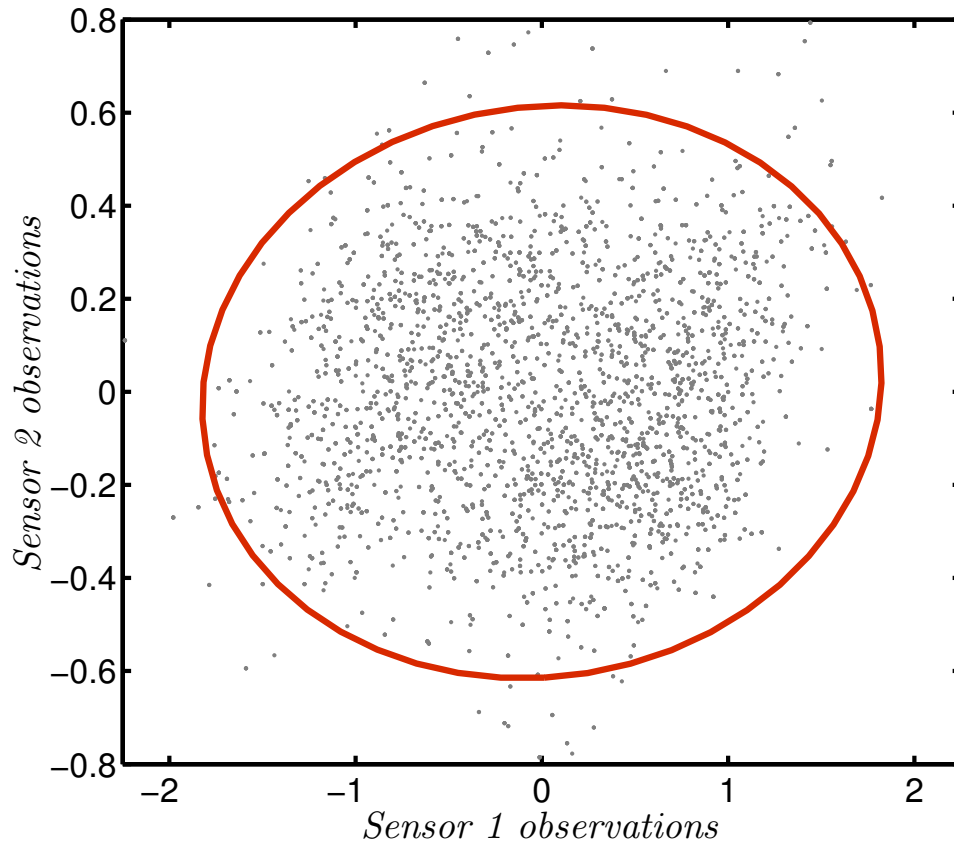


Fig. 4.5: Scatter plot of pilot background data (5000 observation pairs). The 99% confidence ellipse is shown for a $\mathcal{N}([0, 0]^T, \Sigma)$ distribution, where $\Sigma_{1,1} = 0.58$, $\Sigma_{1,2} = \Sigma_{2,1} = 0.009$ and $\Sigma_{2,2} = 0.05$ are the elements of the covariance matrix Σ .

Table 4.3: Summary statistics for parameter estimates of seismic data: Mean, Median and [Standard Error, Median Absolute Deviation]

<i>Hyp. & Sensor i</i>		α	β	γ	δ	τ
BG	1	2, 2 [0, 0]	- [-]	0.54, 0.54 [0.001, 0.03]	$5.1 \cdot 10^{-4}$, 0 [$4.8 \cdot 10^{-4}$, 0.02]	0.02, 0.03 [0.002, 0.06]
	2	2, 2 [0.001, 0]	- [-]	0.17, 0.17 [0.0005, 0.01]	$6 \cdot 10^{-4}$, 0 [$3 \cdot 10^{-4}$, 0.01]	
FS	1	1.28, 1.23 [0.01, 0.32]	0.05, 0.01 [0.007, 0.039]	0.49, 0.47 [0.004, 0.1]	0.02, 0.01 [0.004, 0.03]	0.1, 0.1 [0.003, 0.06]
	2	1.19, 1.14 [0.01, 0.28]	-0.004, 0 [0.005, 0.04]	0.51, 0.48 [0.005, 0.11]	0.002, 0 [0.0003, 0.03]	

BG: Background; FS: Footstep.

Median Absolute Deviation, $\text{MAD}(Z) = \text{median}(|z_j - \text{median}(Z)|)$

elements are obtained by noting that the correlation coefficient $\rho = \sin(\tau\pi/2)$ for a Gaussian copula. A tail-index value of $\alpha = 2$ and the scatter plot in Fig 4.5 suggest that a bivariate normal is a satisfactory model for the background data. Note that a bivariate Gaussian copula with Gaussian marginals is, in effect, a bivariate Gaussian.

For the footstep data, the MAD for β and δ values is significantly smaller than the MAD for α and γ values. The footstep data are, therefore, modeled as $S\alpha S$ distributions, with unknown α and γ . Histogram plots for α and γ corresponding to footstep signals from each sensor are shown in Fig. 4.6. We observe that the values of α and γ do not cluster around the median values reported in Table 4.3. Instead, they show approximately bimodal behavior. We infer that this behavior is observed because footfalls are separated by periods of interstitial background.

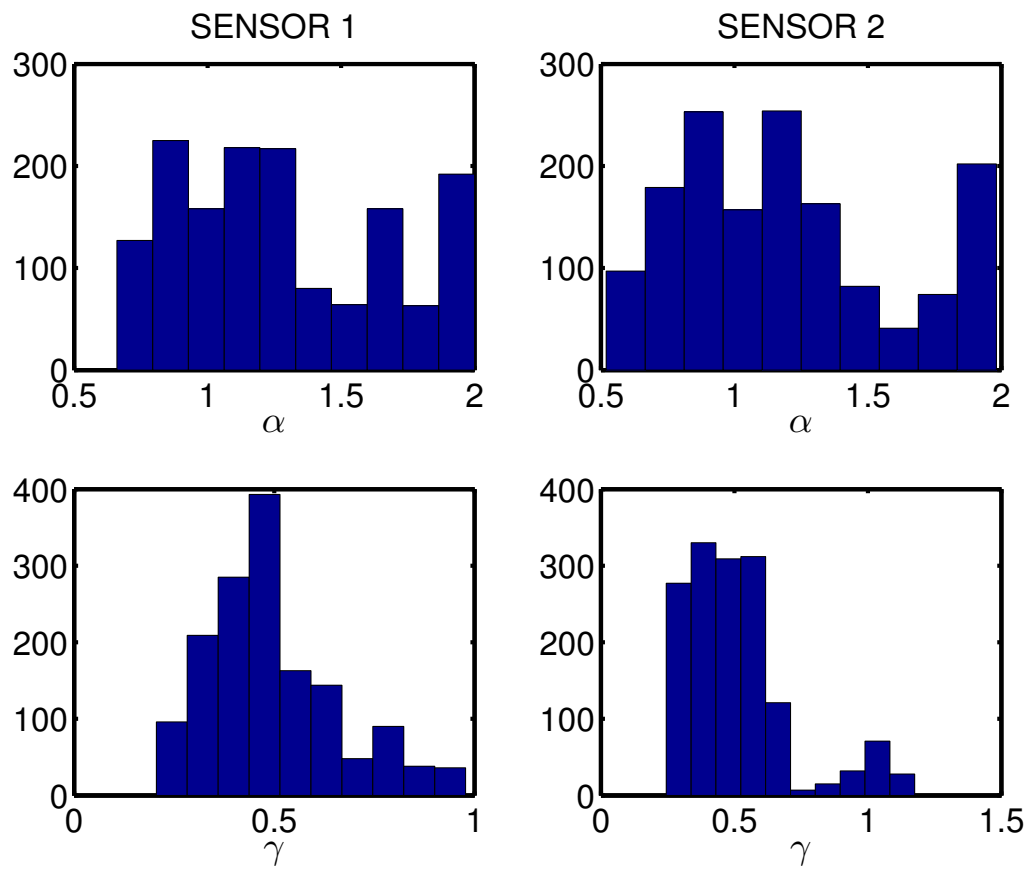


Fig. 4.6: Histogram of α and γ values for the footsteps data.

Results

The detection problem for the background hypothesis, H_0 , vs. the footsteps hypothesis, H_1 , is formulated as,

$$\begin{aligned}
 H_0 : \quad & [X_1, X_2]^T \sim \mathcal{N}([0, 0]^T, \Sigma) \\
 H_1 : \quad & X_1 \sim \mathcal{S}(\boxed{\alpha_1}, 0, \boxed{\gamma_1}, 0) \quad X_2 \sim \mathcal{S}(\boxed{\alpha_2}, 0, \boxed{\gamma_2}, 0) \\
 & \boxed{c_1}
 \end{aligned} \tag{4.46}$$

H_0 , as written in (4.46), is the normal representation of the stable distribution as parametrized for background data in Table 4.3, with c_0 being the Gaussian copula. The ROC for the copula-based detector using KRE for marginal parameter estimation is shown in Fig. 4.7. Since, the sensors are of the same modality (geophones), the ROC for individual sensors is also shown for comparison. The copula based fusion, along with the α -stable modeling, yields superior P_D , especially at lower P_F values. Table 4.4 shows the percentage of decision windows selected for each copula family. Recall that the true copula is not known under H_1 . The copula selection process mostly selects the t copula. This is consistent with the tail dependence properties of the t copula, i.e., for the footfall periods a copula that can adequately model co-occurring footstep spikes is chosen. The Frank copula, on the other hand, is an Archimedean copula which has 0 tail dependence [65]. The 40 of 1500 windows ($\approx 2.66\%$), modeled as a Frank copula, is indicative of the lack of tail dependence during the background periods which are interspersed between consecutive footfalls. The copula selection under H_0 is more evenly spread out over the copula library. This is largely due to low τ values, which means that $c_m \approx 1$ for all $c_m \in \mathcal{C}$.

Table 4.4: Percentage of copulas selected under each hypothesis

<i>Hypothesis</i>	Gaussian	Student- <i>t</i>	Clayton	Frank	Gumbel
H_0	27.96	23.04	11.72	23.97	13.31
H_1	0	96.47	0	2.66	0.87

4.5 Summary

In this chapter, we have developed the asymptotic theory for detection performance when sensor data are heavy-tailed and spatially dependent. The spatial dependence was modeled using copula theory, and the heavy-tailed nature of the data were modeled using α -stable data. A detection problem was formulated, in the Neyman-Pearson framework, to discriminate between a known null process and an unknown composite hypothesis. When applied to footstep data, we observed that tail dependence plays an important role in the copulas that are selected as a part of the detection problem. We also observed that, while the α -stable models are able to explain the behavior of the observed data, parameter estimation – necessary for the construction of the test statistic – is a computationally challenging task. The computational load incurred is rather severe, in spite of limiting our simulations and tests on real data to two sensors. In order to scale a copula-based detection scheme for dependent heavy-tailed data, we need to use more tractable and flexible models for both, marginal and copula modeling. One of the alternative approaches to marginal modeling is to use a nonparametric kernel based approach. This was used in the bivariate context on outdoor seismic-acoustic data, collected by ARL, with similar results [34]. This, and other approaches, are discussed in the context of multivariate modeling in Chapter 5.

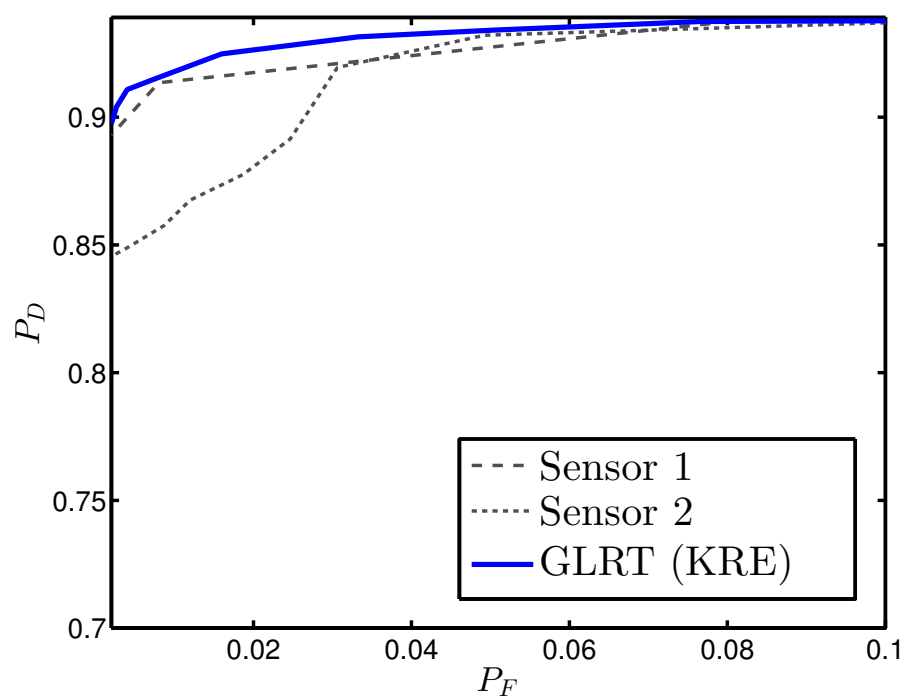


Fig. 4.7: ROC for Background vs. Footstep detection.

CHAPTER 5

DEPENDENCE MODELING FOR DETECTION USING MULTIPLE SENSORS

In Chapter 4, we discussed copula-based detection for heavy-tailed signals and applied the GLR test-statistic to a bivariate or two sensor formulation for copula-based detection. In this chapter, we address copula construction and model selection issues for the multivariate (multisensor) case. These considerations influence the detector design, which will be discussed in detail along with alternative approaches to model the distribution of sensor observations.

The rest of the chapter is organized as follows. The mathematical formulation of the detection problem is presented in Section 5.1. Since we consider the dependence between data acquired by multiple sensors, we need to consider the practical implications of building a multivariate distribution. We elaborate upon the issues pertinent to the construction of a multivariate copula, and thus a multivariate distribution, in Section 5.2. We present our results in Section 5.4 and provide concluding remarks in Section 5.5.

5.1 Problem Formulation

As in Chapter 4, we formulate the detection problem under the Neyman-Pearson framework. We denote the sensor observations by x_{ij} , where $i = 1, 2, \dots, L$ denotes the sensor index and $j = 1, 2, \dots, N$ denotes the time index. That is, a decision window of N samples per sensor is used. Similar to Chapter 4, we make a simplifying assumption that the signals are i.i.d. over time. Also recall from the background data analysis, in Section 4.4.3, that we observed a very low correlation under H_0 . While we explored the detection problem in its full generality in Chapter 4, in this chapter we make a simplifying assumption that the sensor observations are normally distributed and are independent under H_0 . This will facilitate a clearer exposition of the multi-sensor aspects, i.e., multivariate copula, of the detection problem.

We, therefore, have the following binary hypothesis testing problem,

$$H_0 : f_{\mathbf{X}}(\mathbf{x}_j) = \prod_{i=1}^L (\sqrt{2\pi}\sigma_i)^{-1} \exp [-x_{ij}^2/(2\sigma_i^2)] \quad (5.1)$$

$$H_1 : f_{\mathbf{X}}(\mathbf{x}_j) = \prod_{i=1}^L f(x_{ij}) \cdot c(F_1(x_{1j}), \dots, F_L(x_{Lj})|\phi)$$

where $\mathbf{x}_j = [x_{1j}, x_{2j}, \dots, x_{Lj}]$ and σ_i is the standard deviation of the background process observed by the i th sensor. Note that the copula density function $c(\cdot|\phi)$ is a *multivariate* copula density function parametrized by the dependence parameter vector ϕ .

In this chapter, we assess the performance of various Minimum Description Length (MDL) based copula selection schemes. MDL-based model selection schemes are heuristics developed over likelihood-based model selection. They include penalty terms, which are functions of parameter dimensionality and sample size, and emphasize model parsimony. Details on how to obtain the appropriate $c(\cdot|\phi)$ are deferred to the next section. Also, in practice, the parameters σ_i and ϕ are generally not known and are estimated using maximum likelihood estimation (MLE). This is different from Chapter 4, where, in order to develop the theory, we assumed a

simple null hypothesis. Here, both H_0 and H_1 are composite.

The test statistic employed is the likelihood ratio given as,

$$T(\mathbf{x}_j) = \frac{f_{\mathbf{x}}(\mathbf{x}_j | H_1)}{f_{\mathbf{x}}(\mathbf{x}_j | H_0)} \quad (5.2)$$

and is evaluated at the fusion center from the received data. For a given copula density, the marginal distribution under H_1 , $f(x_{ij})$, will determine the performance of the detector. For nonstationary environments, determining the marginal distribution of sensor observations is often a challenging task. We now discuss three possible detection schemes, that essentially model the marginals in different ways. Each scheme gives us a test statistic, $T_k(\mathbf{x}_j)$, $k = 1, 2, 3$. The test compares $\sum_j \log T_k(\mathbf{x}_j)$ against a threshold η ,

$$\sum_{j=1}^N \log T_k(\mathbf{x}_j) \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad k = 1, 2, 3 \quad (5.3)$$

5.1.1 Complete ignorance

In this case, we assume the worst case in that modeling the marginals is not feasible. The detector ignores the marginal information and is completely based on the copula density. The test statistic in this case is,

$$T_1(\mathbf{x}_j) = c(\hat{F}_1(x_{1j}), \dots, \hat{F}_L(x_{Lj}) | \hat{\phi}) \quad (5.4)$$

where $\hat{\phi}$ is the ML estimate of the copula parameters and $\hat{F}_i(x_{ij})$ denotes the empirical probability integral transform for the i -th sensor, which is calculated as in Eq. (4.32). Note that Eq. (5.4) is a likelihood ratio with $f_{\mathbf{x}}(\mathbf{x}_j | H_0) = 1$ since it is the L -fold product of the uniform probability density $\mathcal{U}(0, 1)$. This approach is similar to the detector used in [89]; the difference here is in the construction and use of multivariate copulas. The empirical probability integral transform (EPIT) provides the uniformly distributed arguments for the copula density. For

notational simplicity we use \hat{u}_{ij} , i.e.,

$$\hat{u}_{ij} = \hat{F}_i(x_{ij}) \quad (5.5)$$

In other words, the likelihood ratio uses *only* the information available after variable transformation, i.e., EPIT.

5.1.2 Approximate modeling

Under this scheme, the marginal density is parametrized by a known p.d.f. that can model some critical properties of the signal under consideration. We emphasize that this is *approximate* modeling since, unlike speech signals which have well-established p.d.f. models, it may not be feasible to accurately model other types of signals with sufficient generality. For example, with the footstep data, the signal characteristics are dependent on the type of floor, the background environment and the nature of the mixing of footsteps and background. For our dataset, we have observed that the logistic distribution provides a workable approximation for the heavier tails due to footfalls and is also able to capture the symmetric nature of the geophone signal.

We denote the approximate marginal by $\tilde{f}(x_{ij})$. The test statistic under this scheme is,

$$T_2(\mathbf{x}_j) = \frac{\prod_{i=1}^L \tilde{f}(x_{ij}) \cdot c(\hat{u}_{1j}, \dots, \hat{u}_{Lj} | \hat{\phi})}{\prod_{i=1}^L (\sqrt{2\pi\hat{\sigma}_i^2})^{-1} \exp \left[-x_{ij}^2 / (2\hat{\sigma}_i^2) \right]} \quad (5.6)$$

We assume the approximate marginal density to be the logistic p.d.f. as it has heavier tails than the Gaussian, but at the same time is more tractable than the α -stable density, especially for parameter estimation. For the case of the footstep data mentioned above, recall from Fig. 4.6, that the α values were rather spread out with one of the significant modes occurring at $\alpha \approx 1.3$. Hence, the Cauchy model would try to fit the footstep data with a heavier tail than it actually possesses, and thus the logistic density is a more appropriate choice. The expression for the

logistic density is given by

$$\tilde{f}(x_{ij}) = \frac{e^{-x_{ij}/s}}{s(1 + e^{-x_{ij}/s})^2}, \quad (5.7)$$

where s is the scale parameter and the location parameter for the logistic density function is standardized to 0. The scale parameter is unknown and in order to evaluate the likelihood function corresponding to the marginal density, we have to estimate s .

5.1.3 Nonparametric marginal estimation

Kernel density estimators [102] provide a smoothed estimate, $\hat{f}(x_{ij})$, of the true density. The test statistic for this scheme is, therefore,

$$T_3(\mathbf{x}_j) = \frac{\prod_{i=1}^L \hat{f}(x_{ij}) \cdot c(\hat{u}_{1j}, \dots, \hat{u}_{Lj} | \hat{\phi})}{\prod_{i=1}^L (\sqrt{2\pi\hat{\sigma}_i^2})^{-1} \exp \left[-x_{ij}^2 / (2\hat{\sigma}_i^2) \right]} \quad (5.8)$$

The choice of bandwidth of a kernel based estimator largely determines the accuracy of the density estimate. The kernel bandwidth is chosen using leave-one-out cross-validation. The selected bandwidth, h^* , is the minimizer of the cross-validation estimator of risk, \hat{J} , for a kernel, K . The risk estimator may be easily computed using the approximation,

$$\hat{J}(h) = \frac{1}{hN^2} \sum_p \sum_q K^* \left(\frac{X_p - X_q}{h} \right) + \frac{2}{Nh} K(0) + \mathcal{O} \left(\frac{1}{N^2} \right), \quad (5.9)$$

where $K^*(x) = K^{(2)}(x) - 2K(x)$ and $K^{(2)}(z) = \int K(z - y)K(y)dy$ (see [102], p. 136). The Gaussian kernel was selected, so that $K(x) = \mathcal{N}(x; 0, 1)$ and $K^{(2)}(z) = \mathcal{N}(z; 0, 2)$. Therefore,

$$h^* = \arg \min_h \hat{J}(h)$$

The complete ignorance, logistic, and nonparametric models provide three possible alternatives for modeling sensor data. These marginal models are combined with the copula models,

discussed next, to form the joint p.d.f.

5.2 Construction of Multivariate Copulas

In this section, we address the issue of constructing the multivariate copula density function. This is an important issue, since although there are a large number of copula functions defined in the literature, the majority of them are defined only as bivariate distributions.

An exception to this is the family of elliptical and Archimedean copulas. For example, Jouini and Clemen [42] discuss the use of Archimedean copulas for aggregating expert opinions from a team of decision makers. A rather severe limitation of using Archimedean copulas, however, is that the experts in the decision making problem are necessarily exchangeable. That is, the experts (sensors) are identical for the decision making task. This is not reasonable when dealing with heterogeneous data. Elliptical copulas place restrictions of symmetry on the nature of dependence, which need not hold true in general. The next subsection discusses a tree-based approach to model multivariate dependence. The method discussed uses a hierarchical, pairwise scheme and is free of symmetry and exchangeability restrictions.

In the discussion that follows (Section 5.2.1), we assume that the copula parameter ϕ is known, and do not include it explicitly in the copula function for brevity of notation. We note that the copula parameter is typically estimated as a part of the copula selection process (Section 5.2.2).

5.2.1 Vines

Kurowicka and Cooke [53] discuss a graphical method of constructing copulas using *vines*. A vine is a nested set of trees, where the edges of the k th tree are the nodes of the $(k + 1)$ th tree, and each tree has a maximum number of edges. The trees are called dependence vines when they are used to encode dependence structures in multivariate distributions. There are several

vine architectures possible; Bedford and Cooke [10] present a graphical model that focuses on pairwise interactions of dependent variables using *regular vines*.

Two types of regular vines have been analyzed in the literature [2] in the context of expressing multivariate copulas: the canonical vines or C-vines, and the drawable-vines or D-vines. For our work, we use the D-vine architecture since they are better suited to our application; the C-vines are useful when it is known that a particular sensor plays a key role in governing inter-sensor dependencies. A vine, regular-vine and D-vine are formally defined below.

Definition 5.1. \mathcal{V} is a vine on K elements if,

1. $\mathcal{V} = (\mathsf{T}_1, \dots, \mathsf{T}_{K-1})$
2. T_1 is a connected tree with nodes $\mathsf{N}_1 = \{1, \dots, K\}$ and edges E_1 ; T_k is a connected tree with nodes $\mathsf{N}_k = \mathsf{E}_{k-1}$ for $k = 2, 3, \dots, K - 1$

\mathcal{V} is a **regular vine** on K elements if it satisfies the additional proximity condition,

3. For $k = 2, \dots, K - 1$, if a and b are nodes of T_k connected by an edge in T_k , where $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are edges in T_{k-1} , then exactly one of a_1, a_2 equals one of b_1, b_2 .

A regular vine is called a **D-vine** if each node in T_1 has a degree of at most 2. A D-vine over 4 elements is shown in Fig. 5.1. When a four-variate joint distribution is defined over this vine, we are essentially establishing a hierarchical, pairwise dependency relation, which can be expressed through copulas. Each tree in the vine represents a decomposition obtained by successively conditioning the variables. We elaborate on this procedure below using the example of a vine over three nodes.

Consider a vine over nodes $\{n_1, n_2, n_3\} = \mathsf{N}_1$. For notational convenience, in this section, we drop the time index j . Each node $n_i \in \mathsf{N}_1$ observes data x_i . We can write the following

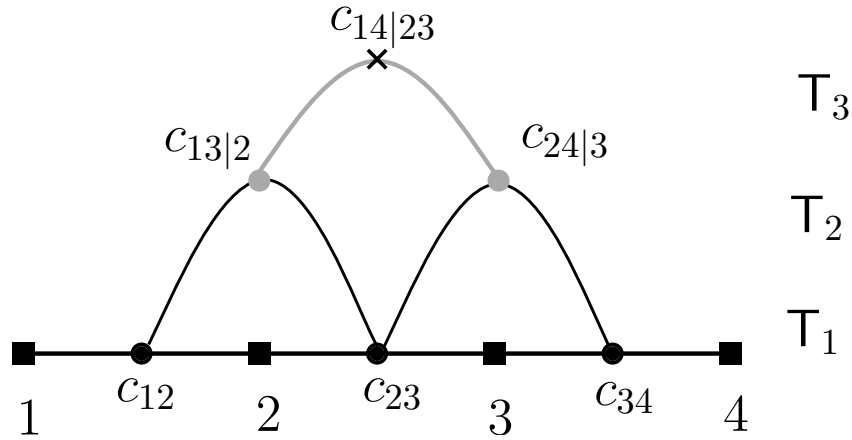


Fig. 5.1: D-vine over 4 elements. Labels indicate the copula density evaluated at each tree in the vine.

pair densities,

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f(x_1)f(x_2) \quad (5.10)$$

$$f(x_2, x_3) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f(x_2)f(x_3) \quad (5.11)$$

where we use subscripts for the copula density function to clarify the node pairs under consideration. In the context of vines, the copula between pairs of nodes is also referred to as the *pair-copula density*. The conditional density for the pair-copulas above is,

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f(x_1) \quad (5.12)$$

$$f(x_3|x_2) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f(x_3) \quad (5.13)$$

From Equations (5.12) and (5.13) we can derive the conditional CDFs that can be used as arguments for the copula defined for tree T_2 in the vine. It is easily seen that,

$$f(x_1, x_3|x_2) = f(x_1|x_2)f(x_3|x_2) \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \quad (5.14)$$

The joint density of x_1, x_2 and x_3 is,

$$\begin{aligned}
f(x_1, x_2, x_3) &= f(x_2)f(x_1|x_2)f(x_3|x_2)c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
&= f(x_1, x_2)f(x_3|x_2)c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
&\stackrel{(a)}{=} f(x_1)f(x_2)c_{12}(F_1(x_1), F_2(x_2))f(x_3|x_2) \\
&\quad \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
&\stackrel{(b)}{=} f(x_1)f(x_2)f(x_3)c_{23}(F_2(x_2), F_3(x_3)) \\
&\quad \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)),
\end{aligned} \tag{5.15}$$

where (a) follows from Eq. (5.10) and (b) follows from Eq. (5.11). In a similar manner, it can be shown that the joint p.d.f. for a 4 variable D-vine is [2],

$$\begin{aligned}
f(x_1, x_2, x_3, x_4) &= f(x_1)f(x_2)f(x_3)f(x_4) \\
&\quad \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4)) \\
&\quad \cdot c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{24|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\
&\quad \cdot c_{14|23}(F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3))
\end{aligned} \tag{5.16}$$

The labels in Fig. 5.1 indicate the copula density evaluated at each tree in the vine. The density of an L -dimensional distribution expressed in terms of a D-vine decomposition is given by Bedford and Cooke [9],

$$\prod_{i=1}^L f(x_i) \prod_{j=1}^{L-1} \prod_{k=1}^{L-j} c_{j,j+k|j+1,\dots,j+k-1}(F(x_j|x_{j+1}, \dots, x_{j+k-1}), F(x_{i+j}|x_{j+1}, \dots, x_{j+k-1})) \tag{5.17}$$

5.2.2 Copula selection

The importance of copula selection has been noted at various points as a vital component of copula-based designs [36, 90]. The dependence between the sensor observations may get

manifested in different ways and the copula function that best models it should be selected. Selecting a copula function that does not adequately model the statistical dependence between the sensor observations may result in model mismatch subsequently deteriorating the detection performance.

When constructing multivariate copulas using vines, the copula selection process has to be repeated for each pair-copula in every tree in the vine. We use a minimum description length (MDL) [33] based approach for model selection. MDL techniques of model selection are based on the principle that the model that achieves the best compression is the model best suited, from the available alternatives, to describe the data. In our case, we do not know the “true” copula density, $c(\cdot)$, and the best possible copula is selected from library of copula functions introduced in Chapter 3.

In this chapter, we compare four criteria available under the MDL framework; the criteria considered are: (1) Akaike Information Criterion (AIC), (2) Bayesian Information Criterion (BIC), (3) Stochastic Information Complexity (SIC) and (4) Normalized Maximum Likelihood (NML). Suppose the copula functions are parametrized by ϕ of dimensionality d . Then, these MDL criteria are defined, for the connected node pairs $\{n_1, n_2\} \in \mathbb{T}_k$, as:

$$AIC = - \sum_{j=1}^N \log c(F(x_{n_1j}), F(x_{n_2j})) + d/2 \quad (5.18)$$

$$BIC = - \sum_{j=1}^N \log c(F(x_{n_1j}), F(x_{n_2j})) + \frac{d}{2} \log N \quad (5.19)$$

$$SIC = - \sum_{j=1}^N \log c(F(x_{n_1j}), F(x_{n_2j})) + \frac{1}{2} \log |\hat{\Sigma}| \quad (5.20)$$

$$\begin{aligned} NML &= - \sum_{j=1}^N \log c(F(x_{n_1j}), F(x_{n_2j})) + \frac{d}{2} \log \left(\frac{N}{2\pi} \right) \\ &\quad + \log \int \sqrt{|I(\phi)|} d\phi \end{aligned} \quad (5.21)$$

where $|\hat{\Sigma}|$ in Eq. (5.20) denotes the determinant of the Hessian of

$$-\sum_{j=1}^N \log c(F(x_{n_1j}), F(x_{n_2j})|\hat{\phi}),$$

and $|I(\phi)|$ in Eq. (5.21) is the determinant of the Fisher information matrix evaluated over $c(F(x_{n_1j}), F(x_{n_2j})|\phi)$.

Copula selection is performed for each bivariate copula term in (5.16). In order to do this, we evaluate AIC , BIC , SIC , and NML for each of the copulas in the copula library (Table 4.2). The copula corresponding to the *minimum* value of each of the MDL criteria is selected. Note that this is similar to the copula selection procedure discussed in Chapter 3: the minimum value is chosen in this case because MDL criteria are defined as functions of the negative log-likelihood.

5.2.3 Node ordering

The D-vine characterization of multivariate dependence constrains the tree T_1 to have a degree of at most 2. This implies that node ordering is important, since different orderings may give rise to different joint-distributions, especially since our copula selection is done through a library of copulae. Therefore, before constructing the D-vine, we must select an appropriate node ordering scheme.

Since our detector capitalizes on the dependency information, we use a dependency criterion to order the nodes. Specifically for the seismic footstep signal detection, we would like to pair sensors that exhibit greater *co-movement* in their signal amplitudes. In other words, we wish to measure the dependence behavior of the sensors at the tails. Chapter 3 discussed the two tail dependency measures called the upper and lower tail dependence. It may be recalled that for a continuous random vector $[X, Y]$, with marginal CDFs F and G , and copula CDF $C(F(X), G(Y))$,

$$\lambda_U \triangleq \lim_{u \nearrow 1} \mathbb{P}(Y > G^{-1}(u) | X > F^{-1}(u)) \quad (5.22)$$

$$= \lim_{u \nearrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \quad (5.23)$$

$$\lambda_L \triangleq \lim_{u \searrow 0} \mathbb{P}(Y \leq G^{-1}(u) | X \leq F^{-1}(u)) \quad (5.24)$$

$$= \lim_{u \searrow 0} \frac{C(u, u)}{u} \quad (5.25)$$

Since a seismic signal oscillates about its mean, λ_U and λ_L provide similar information for characterizing the co-movement of the signal under consideration. In the node ordering algorithm described below we use Eq. (5.23) as the measure by which suitable nodes are paired.

Recall that L denotes the number of sensors and that T_1 is the tree connecting the sensor nodes. Prior to the D-vine construction, we order the nodes in T_1 , $N_1 = \{n_1, n_2, \dots, n_L\}$, for each frame. For this purpose, we implement the following algorithm,

1. Choose a random ordering of sensor nodes. This is only to initialize the ordering process; it is done just once and not repeated for each frame. Label this initial ordering of L nodes as $n_1^1, n_2^1, \dots, n_L^1$. The superscript indexes the iterations through which the algorithm runs.
2. For $l = 1, 2, \dots, L - 2$:

(a) Pick n_l^l and compute λ_U (Eq. (5.23)) for each of the $L - l$ pairs,

$$\{n_l^l, n_{l+1}^l\}, \{n_l^l, n_{l+2}^l\}, \dots, \{n_l^l, n_L^l\},$$

using the copula selected (Section 5.2.2) for that node pair.

- (b) Choose the pairing that has the maximum value of λ_U . Suppose this maximum occurs for some node $k, l \leq k \leq L, k \in \mathbb{Z}^+$, that is, $\{n_l^l, n_k^l\}$ has the greatest value of λ_U among all node pairs. We swap and relabel the nodes as follows:

$$\begin{aligned} n_k^l &\rightarrow n_{l+1}^{l+1} \\ n_{l+1}^l &\rightarrow n_k^{l+1} \\ n_p^l &\rightarrow n_p^{l+1}, \quad p = l+2, l+3, \dots, L \\ &p \neq k \end{aligned}$$

3. The set of nodes $\{n_1^{L-2}, \dots, n_L^{L-2}\} = \mathbf{N}_1$ is the set of nodes in \mathbf{T}_1 ordered by decreasing tail dependence.

5.3 Detection algorithm

In the earlier discussion, we discussed individual components of the detection process. In this section, we list the steps required for the detection detecting an incoming data sequence of length N data from L sensors. We assume that a desired MDL criterion is selected and is used for copula selection. During copula selection, we also assume that ML estimates are used to determine the unknown copula parameters.

1. For each i -th sequence of length N , corresponding to the sensor i , evaluate the marginal likelihood using either logistic or nonparametric methods. For the logistic model, the parameter s in (5.7) is estimated using MLE.
2. Obtain a node order for the base tree \mathbf{T}_1 (Section 5.2.3).
3. Evaluate the conditional CDF arguments for the copulas of \mathbf{T}_2 , as specified in (5.16).
4. Use the copula selection procedure (using the desired MDL criterion) to identify the functions for $c_{13|2}$ and $c_{24|3}$.

5. Using the copula functions for $c_{13|2}$ and $c_{24|3}$, obtain the conditional CDFs for the copula arguments of tree T_3 .
6. Find the copula function, from the copula library, which minimizes the MDL criterion for $c_{14|23}$.

Steps 1-6 give us the log-likelihood under H_1

7. Estimate the variance of the signal, assuming normality (likelihood under H_0)
8. Calculate the test statistic T_k using either of the definitions (5.4), (5.6) or (5.8).
9. Compute the test statistic with a threshold obtained based on the Neyman-Pearson criterion to arrive at the H_0 vs. H_1 decision.

5.4 Results

In this section, we describe the results obtained from various cases considered. The multivariate copula-based detectors, described in the previous section, were applied to the footstep data (Section 2.2). The footstep and background signals are split into non-overlapping frames of 50 samples per sensor. Although the setup consisted of a linear array of 6 sensors, the results presented here use $L = 4$ sensors; the “center” sensors, i.e., 2nd to 5th sensors are used. In the discussion that follows, we use receiver operating characteristics (ROCs) to characterize the detection performance; the ROCs shown are the averages over randomly chosen ensembles of 10 trials. Probability of false alarm and probability of detection are denoted by P_F and P_D , respectively. P_F and P_D are determined empirically by varying the threshold η .

Fig. 5.2 shows the ROCs comparing the different selection criteria discussed in Section 5.2.2. The different selection criteria are compared using $T_1(\mathbf{x}_j)$ (complete ignorance case) since modeling of the marginal distributions has no effect on the copula selection. We observe a slight improvement in overall detection performance with NML as the selection criterion.

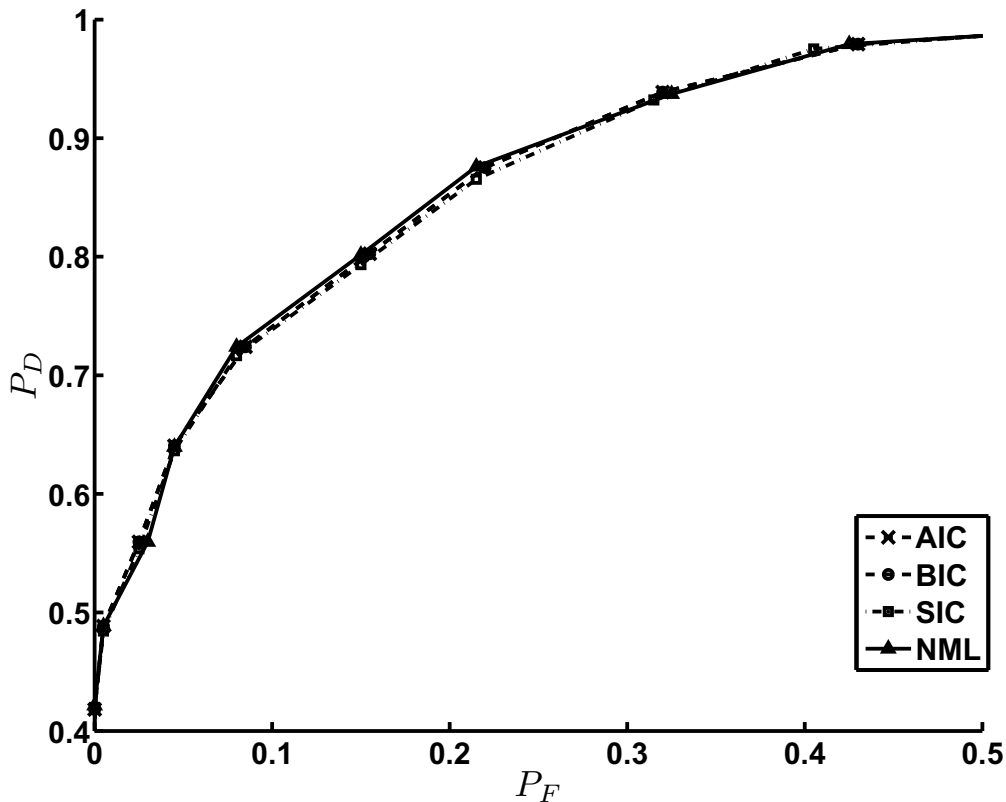


Fig. 5.2: ROC comparing the different MDL-based selection criteria.

However, at lower P_F values SIC is observed to perform better. Note that these are empirical observations made on the footstep data. As such, it has been reported that SIC and NML are superior approaches, especially when used with nested models [33]. We expect that these methods will perform better than AIC and BIC when we consider, e.g., a mixture of copulae. This is a topic for future investigation. NML also incurs a large computational cost because $\int |I(\phi)|$, in Eq. (5.21), is evaluated over all possible values of ϕ . For the remainder of the section, SIC is used as the criterion for copula selection.

Using SIC for model selection from the copula library, we compare the performance of the detectors based on $T_k(\mathbf{x}_j)$ for each $k = 1, 2, 3$. Here we observe, in Fig 5.3, as expected, that the non-parametric modeling of the marginal distribution gives the best detection performance.

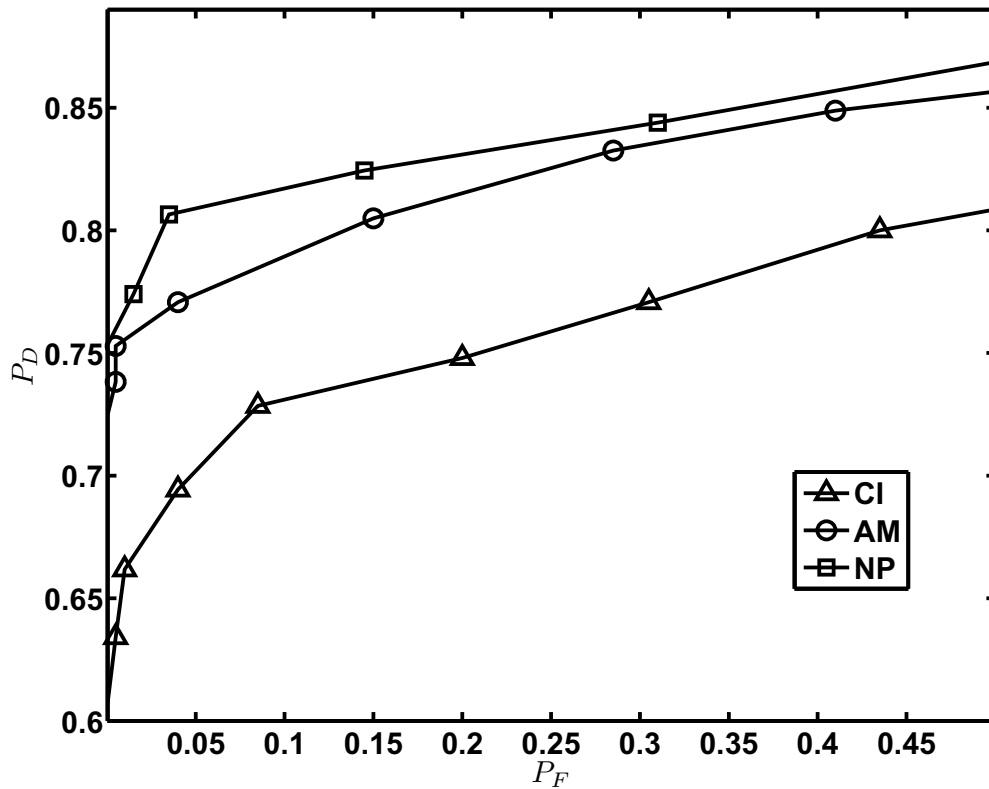


Fig. 5.3: ROC comparing the different detectors obtained from different marginal models. SIC is used for copula selection. CI: Complete Ignorance, AM: Approximate Modeling, NP: Nonparametric modeling.

The approximate model, while easier to compute online, has a lower performance. Detection by ignoring the marginal information also, expectedly, performs poorly in comparison.

One of the key components of the multivariate copula construction was node ordering. The nodes in T_1 correspond to sensor nodes and are ordered using the tail-dependence criterion, λ_U (cf. Eq (5.23)). Recall from Section 2.2 that, in our test-bed setup, sensors are placed as a linear array along a hallway. This suggests that there exists a natural ordering; the sensor closest to one end of the hallway can be sensor 1 and neighboring sensors can be successively indexed as 2, 3 and 4. However, we observe that using λ_U to order the sensors results in a different ordering. This is most likely due to the non-homogeneity of the seismic medium, in this

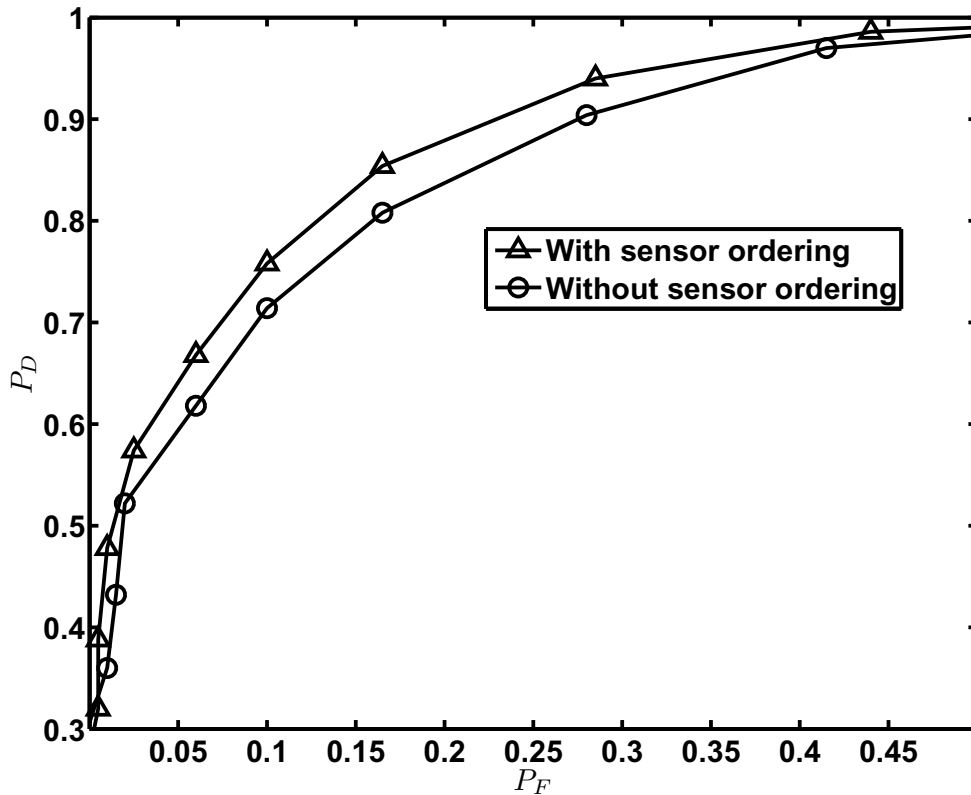


Fig. 5.4: ROC illustrating the benefit of using λ_U -based node ordering.

case, the hallway floor. The natural ordering is used to initialize the node ordering algorithm. The D-vine built using this natural ordering is labeled as the “Without sensor ordering” case in Fig. 5.4. The curve labeled as “With sensor ordering” is the detection performance corresponding to the D-vine built using the end result of the node ordering algorithm. We observe that tail-dependence based ordering gives superior detection performance. While the node order changes for each data frame, i.e., each block of N samples, we observe some consistencies in the pairing patterns. We observe $\{2, 4\}$ and $\{3, 5\}$ are consistently paired together. This is also consistent with the nature of the time-series behavior observed in Fig. 2.4, where, for a given time-interval, footstep spikes occur over the same scale for these sensor pairs.

Constructing a multivariate copula leads to increased system complexity as well as addi-

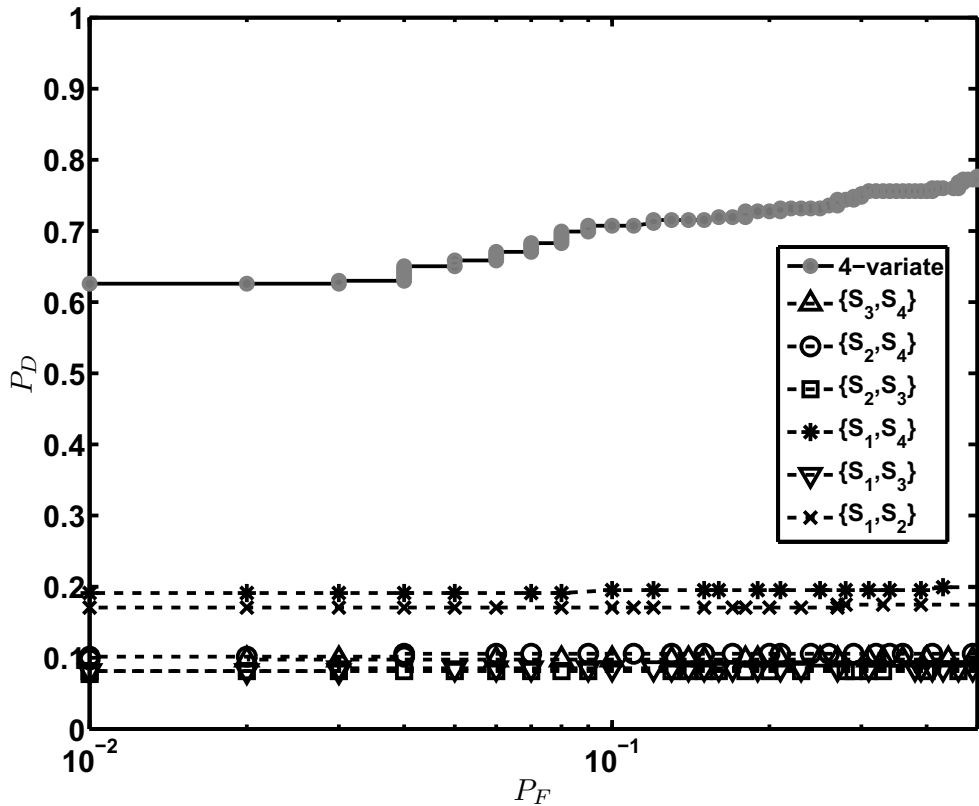


Fig. 5.5: ROC comparing multivariate copula based detector to various bivariate copula based detectors. $\{S_p, S_q\}$ represent sensor pairs for $p, q = 1, 2, 3, 4$ and $p \neq q$. The x-axis is on a logarithmic scale to emphasize low P_F values.

tional computational effort. Fig. 5.5 shows that the additional complexity leads to substantial gains in detection performance when compared to bivariate copula based detectors. In Fig. 5.5, the detection results are presented for the 4-variate case versus results for a number of bivariate cases, considering various pairs of sensors. Since the comparative performance depends on the copula function alone, the ROCs are obtained from detectors that ignore the marginal information.

5.5 Summary

In this chapter, we have discussed detection schemes that consider multisensor dependence using a copula-based approach. Our detector, designed in the Neyman-Pearson framework, demonstrates that accounting for multivariate dependence leads to significant improvement over a bivariate approach. The vine-based approach is suitable for modeling asymmetric dependencies in a tractable manner. However, the computation of conditional CDFs and densities, when repeated for many nodes, leads to some computational burden. Stream based parallel computing solutions are attractive to solve such inference problems, when information fusion is required over large scale networks. These issues, such as computational problems and distribution estimation in a decentralized framework, are open problems ripe for future research in the field of data fusion. The next chapter concludes this dissertation by summarizing the main contributions discussed in this and previous chapters, and also discusses several ideas for future research.

CHAPTER 6

DETECTION OF FOOTSTEPS FROM OUTDOOR DATA

In this chapter, we evaluate the performance of the copula-based detector on seismic and acoustic data collected by the U.S. Army Research Laboratory at the southwest US border. While the previous chapters considered fusion of seismic data from an indoor environment, this chapter considers the fusion of seismic and acoustic data from an *outdoor* environment. The next sections describe the data-collection and the detection performance.

6.1 Data collection and preprocessing

We used the footstep data, made available by the US Army Research Laboratory (ARL), collected at the southwest US border. The dataset consists of raw observations from several sensors of different modalities that were deployed in an outdoor space to record human and animal activity that is typical in perimeter and border surveillance scenarios. The participants in the data collection exercise walked/ran along a predetermined path with sensors laid out along either side of the path.

Seismic and acoustic time series for activities representing a single person walking, two persons walking and human leading an animal (among other examples) are available in the ARL dataset. Each seismic/acoustic time series contains a leading 60s of background data. We use this as our H_0 data. The data are sampled at 10kHz, and are mean centered and oscillatory in nature.

Before applying the copula-based detector, we first pre-process the data. The time series is split into non-overlapping frames of length $T = 512$. This raw time series data is called $x_{Ti}(t)$ where $i = 1, 2$ is the sensor index for the acoustic and seismic modalities respectively, and t is the time index. In keeping with Houston's analysis that Fourier spectra for seismic and acoustic footstep data are more informative than time-domain measurements [35], we set

$$\tilde{x}_{ij} = \sqrt{\mathcal{F}\{x_{Ti}(t)\}^2},$$

and

$$x_{ij} = \tilde{x}_{ij} - \frac{1}{N} \sum_j \tilde{x}_{ij},$$

where \mathcal{F} is the DFT and $j = 1, \dots, N = 256$ is the frequency index. Our sensor measurements are, therefore, now transformed to the frequency domain and the statistics of $\mathbf{x} = [x_{ij}]$ are used as the input to the detector.

Under the background hypothesis we have observed that x_{ij} are normally distributed and are spatially independent. For the footstep hypothesis, we use a copula library consisting of Gaussian, Gumbel and Frank copulas. We have observed that due to the interstitial nature of footstep data, including the independence copula (Table 4.2) in the library improves the overall detection performance.

6.2 Overview of the detector

In this chapter, the hypotheses are formulated as follows:

$$\begin{aligned} H_0 : f_0(\mathbf{x}) &= \prod_{j=1}^N \left[\left(\prod_{i=1}^L f_0(x_{ij} | \boldsymbol{\theta}_{0i}) \right) \right] \\ H_1 : f_1(\mathbf{x}) &= \prod_{j=1}^N \left[\left(\prod_{i=1}^L f_1(x_{ij} | \boldsymbol{\theta}_{1i}) \right) \right. \\ &\quad \left. \times c_1(u_{1j}^1(\boldsymbol{\theta}_{11}), \dots, u_{Lj}^1(\boldsymbol{\theta}_{1L}) | \boldsymbol{\phi}_1) \right]. \end{aligned} \quad (6.1)$$

where $L = 2$, and the index i represents a seismic sensor and an acoustic sensor. Here the distribution under the null hypothesis of background, f_0 , is a the Gaussian p.d.f. For this application, establishing a stationary model under H_1 is not feasible. Therefore, f_1 is determined non-parametrically and u_{ij} is obtained using the empirical probability integral transform (EPIT). The test statistic is, therefore, expressed as,

$$\begin{aligned} T_{\text{NPM}}(\mathbf{x}) &= \sum_{j=1}^N \sum_{i=1}^L \log \frac{\hat{f}_1(x_{ij})}{f_0(x_{ij} | \hat{\boldsymbol{\theta}}_{0i})} \\ &\quad + \sum_{j=1}^N \log c^*(\hat{u}_{1j}, \dots, \hat{u}_{Lj} | \hat{\boldsymbol{\phi}}_{c^*}), \end{aligned} \quad (6.2)$$

where c^* is obtained using the copula selection process described in Chapter 4. The uniform random variables in the copula density are evaluated using EPIT,

$$\hat{F}_i(\cdot) = \frac{1}{N} \sum_{j=1}^N \mathbb{I}_{x_{ij} < \cdot} \quad (6.3)$$

$$\hat{u}_{ij} = \hat{F}_i(x_{ij}) \quad (6.4)$$

where \mathbb{I} is the indicator function. The test is, therefore,

$$T_{\text{NPM}}(\mathbf{x}) \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (6.5)$$

The marginal model under H_1 is determined through a kernel density estimation procedure, as described in Chapter 5. Recall that kernel density estimators [102] provide a smoothed estimate, $\hat{f}_1(x_{ij})$, of the true density.

6.3 Results

For the ARL dataset, we use the statistic $T_{\text{NPM}}(\mathbf{x})$ in (6.2). To generate ROCs, we compare the test-statistic to a vector of thresholds. The curve thus generated, for the case when H_1 corresponds to one person walking, is shown in Fig. 6.1. This curve is compared to the ROC for the product rule, i.e., independence assumption for H_1 . Similar ROCs are obtained for the cases of two persons walking, and man leading an animal and are shown in Fig. 6.2 and Fig. 6.3.

For all the three cases, we observe that our proposed method, using copula-based dependence characterization along with copula selection, outperforms the ROC corresponding independence. We further observe that, the two-persons and man-leading-animal cases have a higher probability of detection (P_D) for a given probability of false alarm (P_F), when compared to the one person case. This is intuitive, since for the two-persons case and man-leading-animal case, we have a higher signal to noise ratio.

6.4 Conclusion

The detection results obtained on the outdoor dataset are similar to those obtained on the indoor dataset. Although the outdoor seismic-acoustic environment is quite different from a typical indoor environment, the copula selection process ensures that the dependence is adequately modeled. Hence, the detection performance is also superior when compared to the independence assumption.

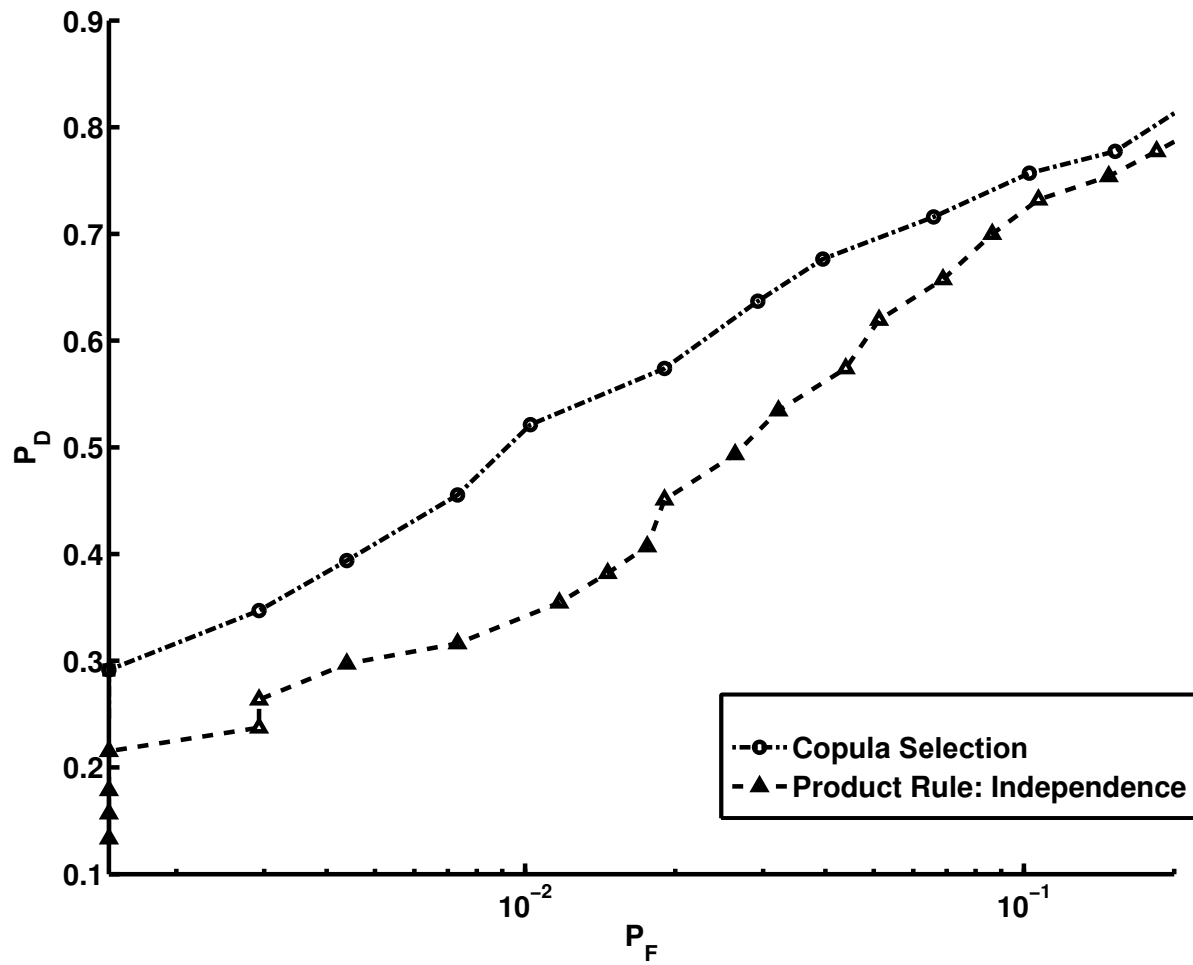


Fig. 6.1: ROCs for the ARL dataset for 1 person vs. background detection.

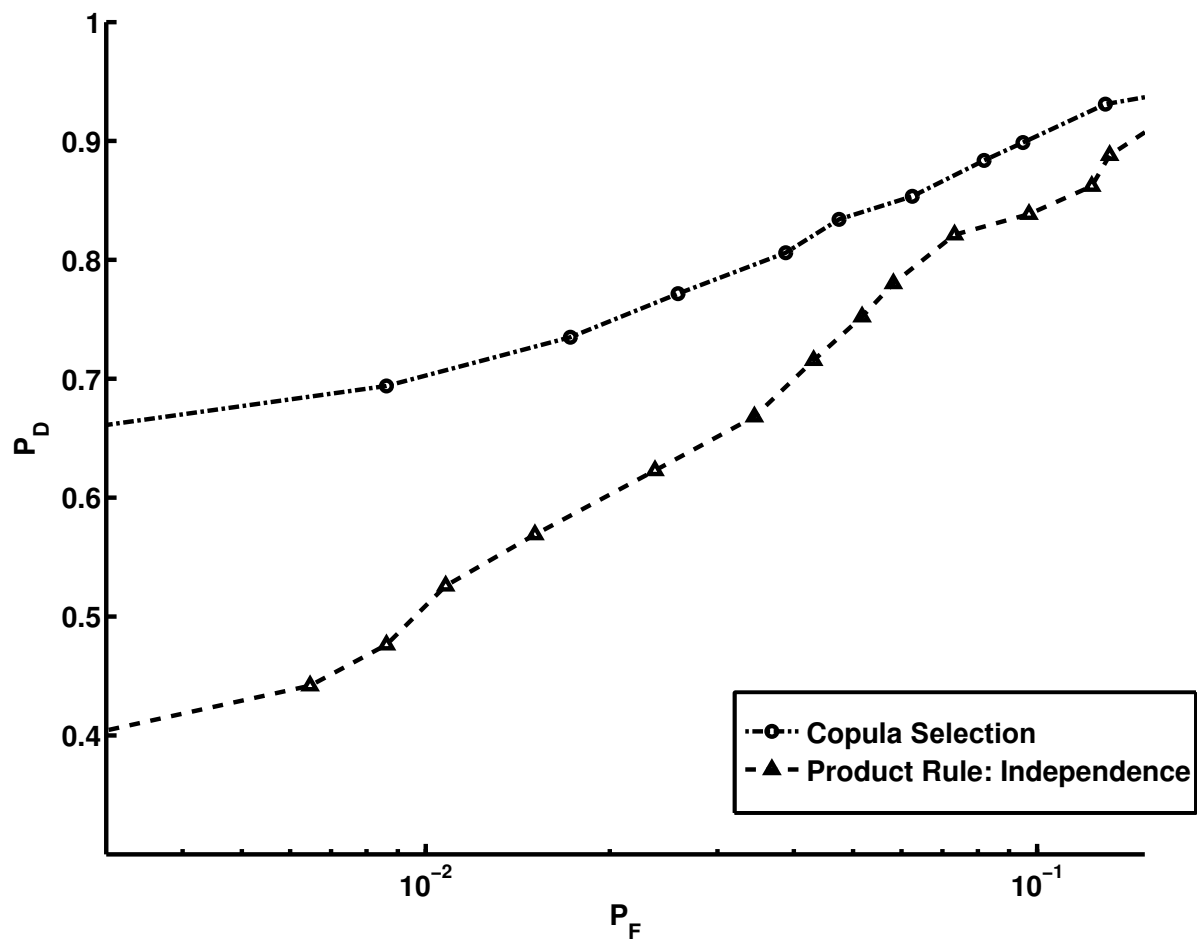


Fig. 6.2: ROCs for the ARL dataset for 2 persons vs. background detection.

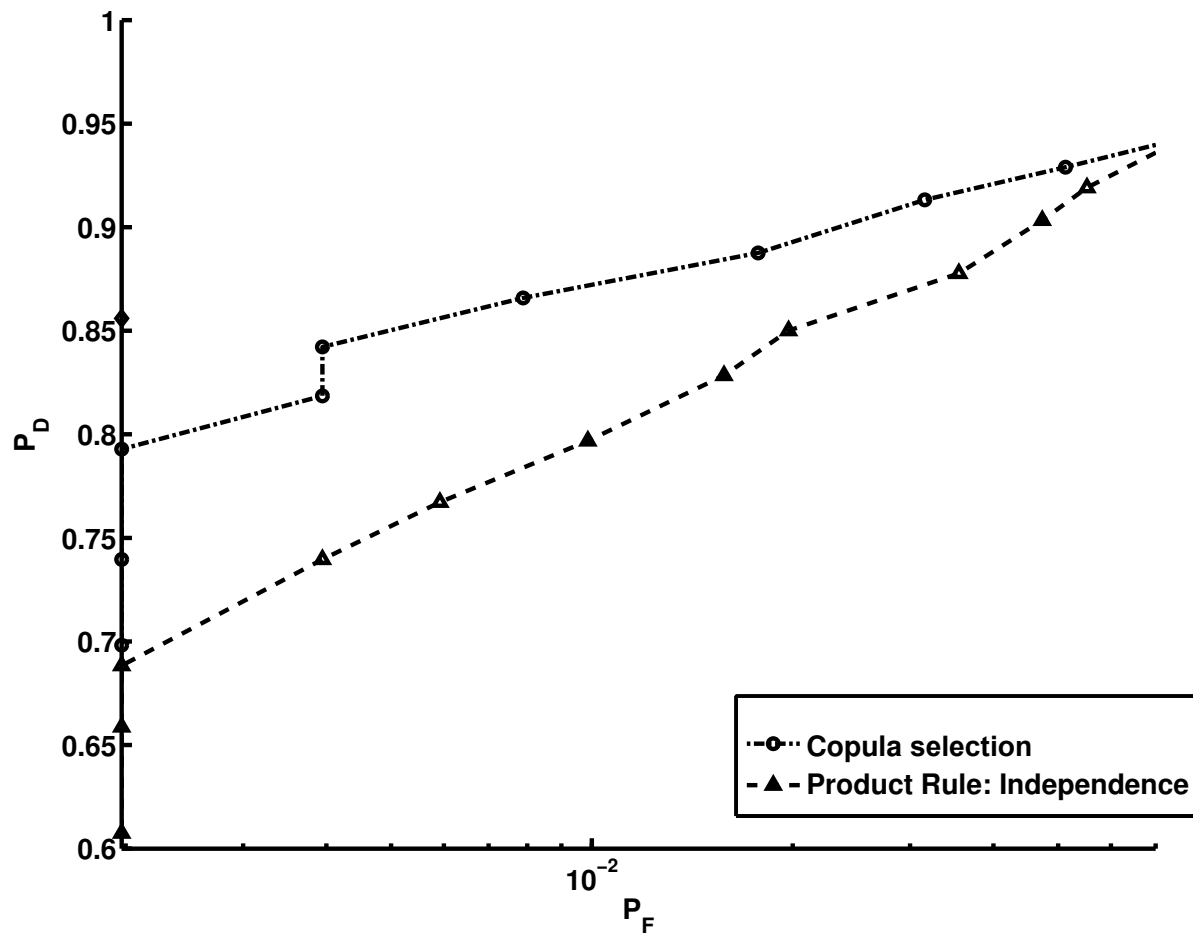


Fig. 6.3: ROCs for the ARL dataset for 1 person leading an animal vs. background detection.

CHAPTER 7

SUMMARY AND FUTURE DIRECTIONS

When monitoring various phenomena, measurements are often nonlinear non-Gaussian and dependent. In this dissertation, we have investigated the design and analysis of detection schemes for one specific class of such data, namely heavy-tailed dependent data. Specifically, we have considered indoor footstep data, obtained using an array of geophone sensors, as a representative example for heavy-tailed, dependent data and have developed appropriate detection schemes.

We used α -stable models as the characterization for heavy-tailed signals. Stable distributions are an important class of models in the study of phenomena where extreme-valued measurements occur with polynomially decaying probability. The co-occurrence of such (rare) extreme-valued data is sometimes symptomatic of a catastrophic event, and its detection, therefore, needs appropriate modeling tools. In this dissertation, we have moved in that direction by addressing the problem of detection for statistically dependent stable distributed heavy-tailed data.

We proposed a copula-based modeling scheme, which allows for a range of possibilities in terms of dependence modeling. It allows for the joint modeling of heavy-tailed marginals with disparate support. It allows us to capture varying degrees of dependence in the tails. The

proposed approach is useful for formulating various detection problems; e.g., it can be used to discriminate between signals embedded in dependent α -stable noise.

We formulated a general two-sided test in the Neyman-Pearson framework, and discussed the need for copula selection. We used a copula selection scheme, which along with likelihood ratio test statistics, ensures asymptotic optimality when the copula corresponding to the data generation process is contained in a copula library. The asymptotic performance of the detectors was studied in detail. We considered several simulated examples, and demonstrated that appropriate models lead to significantly superior detection probabilities, P_D , especially in the low P_F regime.

In order to test our methodology on actual sensor data, we applied the proposed model and detection scheme to the problem of indoor footstep detection. Using the dataset described in Chapter 2, we showed that the footstep data are modeled well using $S\alpha S$ models. Detector and copula selection performance was evaluated. Without significant preprocessing, we obtain $P_D > 0.9$ in the region $P_F \geq 0.01$. The copula functions selected are also consistent with the observed nature of footstep signals in terms of tail dependence behavior.

Finally, we considered the issue of modeling multivariate (i.e., multisensor) dependence. While the theory of copulas can characterize multivariate statistical dependence in full generality, most copula functions, which have been proposed in the literature, are effective only for bivariate dependence characterization. We used a vine-based approach to express the multivariate copula density, in which observation pairs from different sensors are combined in an hierarchical manner. Additionally, we also considered different approaches to modeling the marginal p.d.f.s so that the computational effort of fitting an α -stable distribution is somewhat mitigated. A GLR test statistic was designed for the footstep detection problem, and the performance of minimum description length (MDL) based copula selection schemes were investigated. An important aspect of multivariate modeling is to ascertain which nodes should be paired at the base tree of the vine-based copula model. We have proposed a tail-dependence

based algorithm for node ordering. Our results on the footstep data show that the proposed scheme yields significant performance improvement when compared to the case where only two sensors are used.

7.1 Future directions

Based on the modeling and detection problems addressed in this dissertation, several directions for future research may be identified. These ideas are discussed in this section.

7.1.1 Distributed detection with α -stable dependent observations

The proposed detection schemes were implemented as centralized schemes. Further research will be necessary to port the copula based approach onto a fully distributed framework. Iyengar et al. [37] were successfully able to address the issue of computational efficiency in a copula-based approach to fusing dependent sensor decisions in a distributed detection setup by injecting a controlled amount of noise to the sensor output. The approach, in [37], is especially significant for signals with stable (marginal) distributions, since their analysis is largely in the characteristic function domain. A joint characteristic function approach, based on copula theory, would be a natural extension for distributed inference using dependent heavy-tailed signals. The implication of using vine-based topologies, in the context of joint characteristic functions, may also be explored.

7.1.2 Sequential detection

Sequential detection procedures as first established by Wald, are essentially a generalization of the Neyman-Pearson formalization. For a desired size and power, a sequential test procedure seeks to determine the smallest number of observations, N , that satisfy the false alarm and power constraints. This number N is, in fact, $N(X)$: it is a function of the random variable

being observed and is called the Wald stopping variable. As with the Neyman-Pearson formulation, the optimal test statistic is of the form of a (log) likelihood ratio. A notable feature is that in the sequential framework one can *update* the test statistic. That is, the test statistic for a sequence of N observations, x^N , can be expressed as a function of the test statistic computed for the previous $N - 1$ observations,

$$T(x^N) = T(x^{N-1}) + \log \frac{f_X^1(x_N | x_{1:N-1})}{f_X^0(x_N | x_{1:N-1})}$$

The functions f_X^1 and f_X^0 correspond to the distributions under the alternative and null hypotheses respectively. This detection problem does not assume that the observation sequence is independent: f_X^1 is a joint density characterized by a copula and its respective marginals. This skeletal formulation can be explored in greater detail. The following associated issues that will need to be addressed in this context:

- **Spatio-temporal dependence modeling.** The sequential detection problem for dependent observations, in its most general form, will account for observations from multiple sensors and time points. Further study is required to identify which models will yield more insight and allow for tractable analysis. For example, one may consider that linear models such as ARMA models, characterizing temporal dependence, may be used in conjunction with a copula-based models for spatial dependence modeling. A copula-based approach to temporal dependence characterization, may also be investigated, however the type of copula to be used will still remain an important question to be addressed. Specifically, a copula selection process for a sequential inference problem may incur computational overheads that may lead to intolerable latencies in the system being designed. These issues will have to be addressed carefully.
- **Effect of dependence on $N(X)$.** It will be of interest to quantify the effect of dependence on the stopping variable. The copula based dependence characterization will allow

for a greater degree of generality in this analysis.

7.1.3 Misspecified marginals.

Robust detection techniques with ϵ -contamination models are well studied under an IID assumption. The theory provides techniques to handle distribution uncertainty. In the case of the copula based detectors, marginal uncertainty or misspecification leads to arguments in the copula term which are not uniform distributed. Designing detectors for such situations will be useful especially for applications with non-stationary conditions, where there is reason to believe that the marginals (sensor models) may be perturbed, however the nature of the dependence remains unchanged.

7.1.4 Bootstrap-based detection for dependent observations

As observed in Chapter 4, the distribution of the test-statistic for a detection problem can be difficult to obtain. Zoubir and Iskander [112] discuss bootstrap techniques for various signal processing techniques. Bootstrap based detectors are attractive for small-sample copula-based detection as they allow for a non-parametric methodology for threshold determination. As an initial approach for handling correlated data, Zoubir and Iskander consider an autoregressive model and discuss resampling techniques for this specific case of dependent observations. Efron and Tibshirani [28] discuss how bootstrap implicitly samples from the empirical probability distribution. Assuming temporal independence, we can resample observations from within each sensor, using existing bootstrap theory. Sampling separately from the appropriate copula density will create a pool of random vectors, whose dependence structure is unaltered. This idea can be used to design a bootstrap based non-parametric scheme for automatic threshold selection under the dependence regime.

7.2 Some additional open problems

The ideas contained in Section 7.1 emerged as research topics motivated by the discussion in Chapter 4 and Chapter 5. There are, however, fundamental problems of theoretical interest that arise as a result of copula-based dependence modeling. These are enumerated below.

1. **System identification.** In attempting to derive the distribution of a copula-based statistic, one soon encounters that the theory of functions of dependent random variables, as parametrized by copula functions, is still largely unexplored. Only recently Cherubini [18] has defined the copula based convolution to derive the distribution of the sum of bounded and unbounded dependent random variables. Given that the behavior of linear time invariant systems for independent inputs is well developed, can a similar theory be developed, using the copula framework, for a system accepting *dependent* inputs?
2. **Performance bounds and dependency measures.** For the copula-based inference, it is known that selecting an incorrect copula can penalize performance. Iyengar et al. [38] note that a copula based model does not necessarily perform better than one assuming independence; the necessity for copula selection arises from this observation. However, can this be quantified in terms of the various concepts and measures of dependence that were discussed in Chapter 3? In other words, it would be of interest to investigate if certain performance guarantees are available based on the different types of dependence such as positive quadrant dependence (PQD) and likelihood ratio dependence (LRD).

REFERENCES

- [1] V. Aalo and R. Viswanathan, “On distributed detection with correlated sensors: two examples,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414–421, 1989.
- [2] K. Aas, C. Czado, A. Frigessi, and H. Bakken, “Pair-copula constructions of multiple dependence,” *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, April 2009.
- [3] R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhauser, 1998.
- [4] M. Anderson, T. Adali, and X.-L. Li, “Joint Blind Source Separation with Multivariate Gaussian Model: Algorithms and Performance Analysis,” *IEEE Transactions on Signal Processing*, 2011, to be published.
- [5] F. J. Anscombe, “Graphs in statistical analysis,” *The American Statistician*, vol. 27, no. 1, pp. 17–21, Feb. 1973.
- [6] G. R. Arce, *Nonlinear signal processing: a statistical approach*. Hoboken, NJ: Wiley-Interscience, 2005.
- [7] N. Balakrishnan and C. D. Lai, *Continuous Bivariate Distributions*, 2nd ed. New York, NY: Springer, 2009.

- [8] M. J. Beal, N. Jojic, and H. Attias, “A graphical model for audiovisual object tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 828–836, 2003.
- [9] T. Bedford and R. Cooke, “Probability density decomposition for conditionally dependent random variables modeled by vines,” *Annals of Mathematics and Artificial Intelligence*, vol. 32, pp. 245–268, 2001.
- [10] T. Bedford and R. M. Cooke, “Vines: A new graphical model for dependent random variables,” *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002.
- [11] B. Bhanu and V. Govindaraju, Eds., *Multibiometrics for Human Identification*. New York, NY: Cambridge University Press, 2011.
- [12] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Norwood, MA: Artech House, 1999.
- [13] A. Briassouli, P. Tsakalides, and A. Stouraitis, “Hidden messages in heavy-tails: DCT-domain watermark detection using alpha-stable models,” *IEEE Transactions on Multimedia*, vol. 7, no. 4, pp. 700–715, 2005.
- [14] T. Butz and J.-P. Thiran, “From error probability to information theoretic (multi-modal) signal processing,” *Signal Processing*, vol. 85, no. 5, pp. 875–902, May 2005.
- [15] M. Çetin, L. Chen, I. Fisher, J. W., A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, “Distributed fusion in sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, 2006.
- [16] H. Chen, B. Chen, and P. Varshney, “A new framework for distributed detection with conditionally dependent observations,” *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1409–1419, 2012.

- [17] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. Wiley, 2004.
- [18] U. Cherubini, S. Mulinacci, and S. Romagnoli, “On the distribution of the (un)bounded sum of random variables,” *Insurance: Mathematics and Economics*, vol. 48, no. 1, pp. 56 – 63, 2011.
- [19] N. M. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun, “Canonical Correlation Analysis for Data Fusion and Group Inferences: Examining applications of medical imaging data.” *IEEE Signal Process Mag*, vol. 27, no. 4, pp. 39–50, June 2010.
- [20] D. R. Cox, “Tests of separate families of hypotheses,” in *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, J. Neyman, Ed., vol. 1. Berkeley, CA: University of California Press, 1961, pp. 105–103.
- [21] —, “Further results on tests of separate families of hypotheses,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 24, no. 2, pp. 406–424, 1962.
- [22] T. Damarla, “Hidden markov model as a framework for situational awareness,” in *Proc. 11th International Conference on Information Fusion*, 2008, pp. 1–7.
- [23] M. Davy and A. Doucet, “Copulas: a new insight into positive time-frequency distributions,” *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 215–218, Jul. 2003.
- [24] P. D. Ditlevsen, “Observation of α -stable noise induced millennial climate changes from an ice-core record,” *Geophysical Research Letters*, vol. 26, no. 10, pp. 1441–1444, 1999.
- [25] E. Drakopoulos and C.-C. Lee, “Optimum multisensor fusion of correlated local decisions,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 593–606, 1991.

- [26] W. H. DuMouchel, "On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution," *The Annals of Statistics*, vol. 1, no. 5, pp. 948–957, Sep. 1973.
- [27] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: a critique," *The Annals of Statistics*, vol. 11, no. 4, pp. 1019–1031, 1983.
- [28] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall, 1994.
- [29] A. Ekimov and J. M. Sabatier, "Vibration and sound signatures of human footsteps in buildings," *Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 762–768, Aug 2006.
- [30] J. Fang and H. Li, "Power Constrained Distributed Estimation With Correlated Sensor Data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, 2009.
- [31] Geo Space, LP. (2011, March) GS-20DX Specifications. Houston, TX. [Online]. Available: <http://www.geospacelp.com/index.php?id=127>
- [32] B. V. Gnedenko and A. N. Kolmogorov, *Limit Theorems for Sums of Independent Random Variables*, Revised English ed. Reading, MA: Addison-Wesley, 1968.
- [33] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 746–774, 2001.
- [34] H. He, A. Subramanian, P. K. Varshney, and T. Damarla, "Fusing heterogeneous data for detection under non-stationary dependence," in *Information Fusion (FUSION), 2012 15th International Conference on*. IEEE, 2012, pp. 1792–1799.

- [35] K. M. Houston and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," in *Unattended Ground Sensor Technologies and Applications V*, E. M. Carapezza, Ed., vol. 5090, no. 1. SPIE, 2003, pp. 162–173.
- [36] S. Iyengar, P. Varshney, and T. Damarla, "A parametric copula based framework for hypotheses testing using heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2308 – 2319, May 2011.
- [37] S. G. Iyengar, R. Niu, and P. K. Varshney, "Fusing Dependent Decisions for Hypothesis Testing With Heterogeneous Sensors," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4888–4897, 2012.
- [38] S. G. Iyengar, P. K. Varshney, and T. Damarla, "A parametric copula based framework for multimodal signal processing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2009*, Apr. 19–24, 2009, pp. 1893–1896.
- [39] S. G. Iyengar, "Decision making with heterogeneous sensors - a copula based approach," Ph.D. dissertation, Syracuse University, Syracuse, NY, August 2011.
- [40] H. Joe, "Relative Entropy Measures of Multivariate Dependence," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 157–164, 1989.
- [41] H. Joe, *Multivariate Models and Multivariate Dependence Concepts*, 1st ed. Boca Raton, FL: Chapman & Hall/CRC, 1997.
- [42] M. N. Jouini and R. T. Clemen, "Copula models for aggregating expert opinions," *Operations Research*, vol. 44, no. 3, pp. 444–457, 1996.
- [43] M. Kafai and B. Bhanu, "Dynamic Bayesian Networks for Vehicle Classification in Video," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 100–109, 2012.

- [44] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 916–920, 1992.
- [45] R. Kapoor, A. Banerjee, G. A. Tsihrintzis, and N. Nandhakumar, "UWB radar detection of targets in foliage using alpha-stable clutter models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 819–834, 1999.
- [46] S. Kar, P. K. Varshney, and H. Chen, "Spatial whitening framework for distributed estimation," in *Proc. 4th IEEE Int Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) Workshop*, 2011, pp. 293–296.
- [47] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, New Jersey: Prentice-Hall Inc., 1998.
- [48] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, December 1971.
- [49] E. Kidron, Y. Y. Schechner, and M. Elad, "Cross-Modal Localization via Sparsity," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1390–1404, 2007.
- [50] I. A. Koutrouvelis, "An iterative procedure for the estimation of the parameters of stable laws," *Communications in Statistics - Simulation and Computation*, vol. 10, no. 1, pp. 17–28, Jan 1981.
- [51] I. A. Koutrouvelis and D. F. Bauer, "Asymptotic distribution of regression type estimators of parameters of stable laws," *Communications in Statistics - Theory and Methods*, vol. 11, no. 23, pp. 2715–2730, Jan 1982.
- [52] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum Energy Decentralized Estimation in a Wireless Sensor Network with Correlated Sensor Noises," *EURASIP Journal on*

- Wireless Communications and Networking*, vol. 2005, no. 4, pp. 473–482, 2005, article ID 919686.
- [53] D. Kurowicka and R. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, 2006.
- [54] E. E. Kuruoğlu, W. J. Fitzgerald, and P. J. W. Rayner, “Near optimal detection of signals in impulsive noise modeled with a symmetric α -stable distribution,” *IEEE Communications Letters*, vol. 2, no. 10, pp. 282–284, 1998.
- [55] C. D. Lai and M. Xie, *Stochastic Ageing and Dependence for Reliability*. New York, NY: Springer, 2006.
- [56] B. L. Lan and M. Toda, “Fluctuations of healthy and unhealthy heartbeat intervals,” *EPL (Europhysics Letters)*, vol. 102, no. 1, p. 18002, 2013.
- [57] E. L. Lehmann, *Elements of Large Sample Theory*. New York, NY: Springer-Verlag, 1999.
- [58] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, “Detection, classification, and tracking of targets,” *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, 2002.
- [59] J. Li and P. Stoica, “MIMO Radar with Colocated Antennas,” *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.
- [60] D. Mari and S. Kotz, *Correlation and Dependence*. Imperial College Press, 2001.
- [61] Mark Veillette. (2014, Jan) Alpha-Stable distributions in MATLAB. [Online]. Available: <http://math.bu.edu/people/mveillett/html/alphastablepub.html>
- [62] E. L. Melnick and A. Tenenbein, “Misspecifications of the Normal Distribution,” *The American Statistician*, vol. 36, no. 4, pp. 372–373, Nov. 1982.

- [63] G. Mercier, G. Moser, and S. Serpico, “Conditional copula for change detection on heterogeneous sar data,” in *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS 2007*, Jul. 23–28, 2007, pp. 2394–2397.
- [64] B. T. Morris and M. M. Trivedi, “A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114–1127, 2008.
- [65] R. Nelsen, *An Introduction to Copulas*, 2nd ed. New York: Springer, 2006.
- [66] R. Niu and P. K. Varshney, “Target location estimation in sensor networks with quantized data,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, 2006.
- [67] J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhäuser, 2013, ch. 1, pp. 3–24, in progress. [Online]. Available: <http://academic2.american.edu/~jpnolan/stable/stable.html>
- [68] ——. (2014, Apr.) Bibliography on stable distributions, processes and related topics. [Online]. Available: <http://academic2.american.edu/~jpnolan/stable/stable.html>
- [69] —, “Multivariate elliptically contoured stable distributions: theory and estimation,” *Computational Statistics*, vol. 28, no. 5, pp. 2067–2089, Oct. 2013.
- [70] H. Ogata, “Estimation for multivariate stable distributions with generalized empirical likelihood,” *Journal of Econometrics*, vol. 172, no. 2, pp. 248–254, February 2012.
- [71] R. Olfati-Saber and N. F. Sandell, “Distributed tracking in sensor networks with limited sensing range,” in *Proc. American Control Conf*, 2008, pp. 3157–3162.
- [72] A. Ozturk, P. Chakravarthi, and D. Weiner, “On determining the radar threshold for non-Gaussian processes from experimental data,” *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1310–1316, Jul 1996.

- [73] J. Park, G. Shevlyakov, and K. Kim, “Maximin Distributed Detection in the Presence of Impulsive Alpha-Stable Noise,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1687–1691, 2011.
- [74] H. Pesaran and M. Weeks, “Non-nested hypothesis testing: an overview,” in *A Companion to Theoretical Econometrics*, B. H. Baltagi, Ed. Malden, MA: Blackwell Publishing, 2003, ch. 13, pp. 279–309.
- [75] G. Peters, S. Sisson, and Y. Fan, “Likelihood-free Bayesian inference for α -stable models,” *Computational Statistics & Data Analysis*, vol. 56, no. 11, pp. 3743–3756, 2012.
- [76] A. Rajan and C. Tepedelenlioğlu, “Diversity Combining over Rayleigh Fading Channels with Symmetric Alpha-Stable Noise,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2968–2976, 2010.
- [77] A. Rényi, “On measures of dependence,” *Acta Mathematica Hungarica*, vol. 10, pp. 441–451, 1959.
- [78] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor Networks-part I: Gaussian case,” *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, 2006.
- [79] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, 1st ed., ser. International Series on Biometrics. New York, NY: Springer, 2006.
- [80] T. S. Saleh, I. Marsland, and M. El Tanany, “Suboptimal Detectors for Alpha-Stable Noise: Simplifying Design and Improving Performance,” *IEEE Transactions on Communications*, vol. 60, no. 10, pp. 2982–2989, 2012.
- [81] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. Boca Raton, FL: Chapman & Hall/CRC, 1994.

- [82] T. Schreiber and A. Schmitz, “Surrogate time series,” *Physica D*, vol. 142, no. 3, pp. 346–382, 2000.
- [83] ———, “Improved surrogate data for nonlinearity tests,” *Physical Review Letters*, vol. 77, no. 4, pp. 635–638, Jul 1996.
- [84] M. Shao and C. L. Nikias, “Signal processing with fractional lower order moments: stable processes and their applications,” *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.
- [85] F. Simmross Wattenberg, J. I. Asensio Pérez, P. Casaseca-de-la Higuera, M. Martín Fernández, I. A. Dimitriadis, and C. Alberola López, “Anomaly Detection in Network Traffic Based on Statistical Inference and alpha-Stable Modeling,” *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, pp. 494–509, 2011.
- [86] M. Slaney and M. Covel, “FaceSync: A Linear Operator for Measuring Synchronization of Video Facial Images and Audio Tracks,” in *Advances in Neural Information Processing Systems (NIPS)*, vol. 13, 2000, pp. 814–820.
- [87] A. Sundaresan and P. K. Varshney, “Location estimation of a random signal source based on correlated sensor observations,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 787–799, 2011.
- [88] A. Sundaresan, P. K. Varshney, and N. S. V. Rao, “Copula-based fusion of correlated decisions,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 454–471, 2011.
- [89] A. Sundaresan, “Detection and location estimation of a random signal source using sensor networks,” Ph.D. dissertation, Syracuse University, Syracuse, NY, December 2010.
- [90] A. Sundaresan, A. Subramanian, P. K. Varshney, and T. Damarla, “A copula-based semi-parametric approach for footstep detection using seismic sensor networks,” in *Multisen-*

- sor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2010*, J. J. Braun, Ed., vol. 7710, no. 1. SPIE, 2010, pp. 77 100C–77 100C–12.
- [91] A. Swami and B. M. Sadler, “On some detection and estimation problems in heavy-tailed noise,” *Signal Processing*, vol. 82, no. 12, pp. 1829–1846, 2002.
- [92] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, Eds., *Wireless Sensor Networks: Signal Processing and Communications*. Wiley, 2007.
- [93] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, “Testing for non-linearity in time series: the method of surrogate data,” in *Conference proceedings on Interpretation of time series from nonlinear mechanical systems*. New York, NY, USA: Elsevier North-Holland, Inc., 1992, pp. 77–94.
- [94] H. L. V. Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. Wiley-Interscience, 2002.
- [95] G. A. Tsihrintzis and C. L. Nikias, “Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process,” *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 904–914, 1995.
- [96] G. A. Tsihrintzis, M. Shao, and C. L. Nikias, “Recent results in applications and processing of α -stable-distributed time series,” *Journal of the Franklin Institute*, vol. 333, no. 4, pp. 467–497, 1996.
- [97] J. Tsitsiklis and M. Athans, “On the complexity of decentralized decision making and detection problems,” *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.
- [98] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

- [99] J. Voit, *The Statistical Mechanics of Financial Markets*, 3rd ed. Berlin: Springer Verlag, 2005.
- [100] Q. H. Vuong, “Likelihood ratio tests for model selection and non-nested hypotheses,” *Econometrica*, vol. 57, no. 2, pp. 307–333, Mar. 1989.
- [101] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large,” *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426–426, Nov. 1943.
- [102] L. Wasserman, *All of Nonparametric Statistics (Springer Texts in Statistics)*. New York: Springer, 2006.
- [103] G. Weiss, “Time-reversibility of linear stochastic processes,” *Journal of Applied Probability*, vol. 12, no. 4, pp. 831–836, Dec. 1975.
- [104] H. White, “Maximum likelihood estimation of misspecified models,” *Econometrica*, vol. 50, no. 1, pp. 1–25, Jan. 1982.
- [105] ———, “Regularity conditions for Cox’s test of non-nested hypotheses,” *Journal of Econometrics*, vol. 19, no. 2, pp. 301–318, Aug. 1982.
- [106] S. S. Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.
- [107] P. Willett, P. F. Swaszek, and R. S. Blum, “The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise,” *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3266–3279, 2000.
- [108] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *ACM Comput. Surv.*, vol. 38, pp. 1–45, December 2006, article 13.

- [109] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.
- [110] X. Zhong, A. B. Premkumar, and A. S. Madhukumar, "Particle Filtering for Acoustic Source Tracking in Impulsive Noise With Alpha-Stable Process," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 589–600, 2013.
- [111] Z. Zhu and T. S. Huang, Eds., *Multimodal Surveillance: Sensors, Algorithms, and Systems*. Norwood, MA: Artech House, 2007.
- [112] A. M. Zoubir and D. R. Iskander, *Bootstrap techniques for signal processing*. Cambridge University Press, Cambridge, U.K., 2004.

VITA

NAME OF AUTHOR: Arun Subramanian

PLACE OF BIRTH: Mumbai (Bombay), Maharashtra, India

DATE OF BIRTH: August 19, 1981

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, Syracuse, NY, USA

Bhavnagar University, Gujarat, India

DEGREES AWARDED:

Master of Science, 2011, Syracuse University, USA

Bachelor of Engineering, 2003, Bhavnagar University, India

PROFESSIONAL EXPERIENCE:

- Graduate Research Assistant, Sensor Fusion Laboratory, Dept. of Electrical Engineering and Computer Science, Syracuse University, August 2006 - present
- Researcher, Syracuse Medical Imaging Research Group, May 2005 -July 2006
- Research Associate, Department of Mechanical Engineering, IIT Delhi, New Delhi, India, January 2004–June 2004